



## ART-RRT: As-Rigid-As-Possible exploration of ligand unbinding pathways

Minh Khoa Nguyen, Léonard Jaillet, Stephane Redon

### ► To cite this version:

Minh Khoa Nguyen, Léonard Jaillet, Stephane Redon. ART-RRT: As-Rigid-As-Possible exploration of ligand unbinding pathways. *Journal of Computational Chemistry*, 2018, 39 (11), pp.665-678. 10.1002/jcc.25132 . hal-01973778

**HAL Id: hal-01973778**

**<https://inria.hal.science/hal-01973778>**

Submitted on 8 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ART-RRT: As-Rigid-As-Possible Exploration of Ligand Unbinding Pathways

Minh Khoa NGUYEN, Léonard JAILLET, Stéphane REDON \*

December 4, 2017

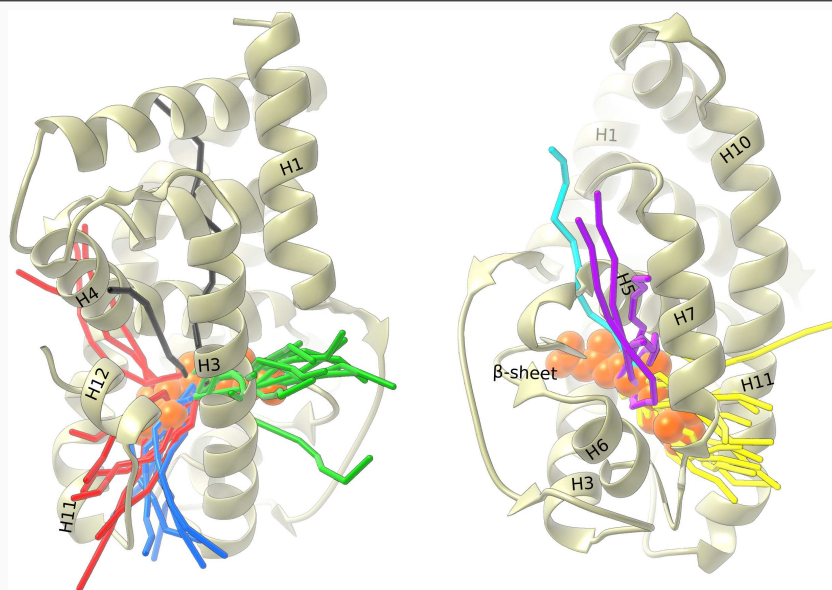
## Abstract

This article proposes a method to efficiently generate approximate ligand unbinding pathways. It combines an efficient tree-based exploration method with a morphing technique from Computer Graphics for dimensionality reduction. This method is computationally cheap and, unlike many existing approaches, does not require a reaction coordinate to guide the search. It can be used for finding pathways with known or unknown directions beforehand. The approach is evaluated on several benchmarks and the obtained solutions are compared with the results from other state-of-the-art approaches. We show that the method is time-efficient and produces pathways in good agreement with other state-of-the-art solutions. These paths can serve as first approximations that can be used, analyzed or improved with more specialized methods.

**Keywords:** ligand unbinding pathway, motion planning, ART-RRT, dimension reduction, morphing. ■

---

\*Université Grenoble Alpes, Inria, CNRS, Laboratoire Jean Kuntzmann, 38000 Grenoble, France



The study of ligand binding/unbinding pathways has become an important subject for protein engineering and computer-aided drug design because it allows the prediction of kinetics properties and identification of the important interactions, which hinder/facilitate the ligand entry/exit. The article proposes a new method, ART-RRT, for generating approximate ligand unbinding pathways in reasonable time. These paths can be used, analyzed or improved with more specialized methods.

# INTRODUCTION

Protein-ligand interactions play important roles in many functions of organisms such as nutritive transport, signaling pathways, hormone regulation, etc. To study these interactions, researchers can take advantage of the Protein Data Bank<sup>1</sup> for protein-ligand binding complexes, advanced docking programs<sup>2</sup> for predicted complexes, or user-friendly web services<sup>3-5</sup> for their interactions. Recently, the study of protein-ligand pathways has become an important research subject since it can validate the predicted complexes from docking applications<sup>6</sup> and allows the prediction of kinetics properties such as binding affinity<sup>7</sup>, binding potency<sup>8</sup> or binding/unbinding rates<sup>9,10</sup>. Pathways can also help identify protein sites which hinder/facilitate the ligand entry/exit<sup>11,12</sup> or whether certain entry/exit routes are more favorable<sup>13-16</sup>. Such information is very valuable for protein engineering and computer-aided drug design<sup>17</sup>.

Despite the increase of computational power, the simulation of ligand binding/unbinding pathways is still a great challenge due to the presence of transition states which are rare, short-lived and associated with high-energy barriers<sup>18</sup>. The timescale of these phenomena usually cannot be probed by classical simulation methods<sup>9</sup>. Therefore, most current methods accelerate classical molecular dynamics (MD)<sup>19</sup> or Monte Carlo (MC)<sup>20</sup> simulations by manipulating temperature, potential energy or force. The approaches that manipulate the temperature include temperature-accelerated MD<sup>21</sup>, replica-exchange MD<sup>22</sup>, simulated annealing<sup>23</sup>, etc. Biased-potential methods such as metadynamics<sup>24</sup>, umbrella sampling<sup>25</sup>, hyperdynamics<sup>26</sup>, local elevation method<sup>27</sup> or conformational flooding<sup>28</sup> favor escaping high-energy barriers and help prevent revisiting them.

For protein-ligand pathways, a common approach is to use biased forces to steer a simulation along certain directions. Among popular methods, one can find Targeted MD (TMD)<sup>29</sup>, Steered MD (SMD)<sup>30</sup>, Random Acceleration MD (RAMD)<sup>31</sup>, etc. In TMD, a biased force based on the Root Mean Square Distance (RMSD) with respect to a certain target structure is added. Hence, the target structure must be known beforehand. With SMD, an external force is applied on a group of atoms along a desired direction. During simulation, this force decreases when the controlled atoms move in the prescribed direction, and increases other-

wise. Hence, SMD is mainly used for analyzing pathways following the directions known in advance. The RAMD method also applies an external force on a group of atoms. However, this force is activated for a certain number of time steps and then changed in direction if the atom group has not moved significantly. Since RAMD is not biased along any specific directions, it can be used for exploring all possible pathways, although the success of the method may be sensitive to the parameter choice<sup>32</sup>. During the past few decades, these methods have contributed significantly to the study of ligand-protein interaction pathways<sup>6,11,16,33–37</sup>.

In recent years, motion planning methods from Robotics have been successfully applied for finding molecular pathways<sup>38,39</sup>. These graph-based methods are known to efficiently explore conformational spaces. The MOMA-LigPath server<sup>40</sup>, which extends the ML-RRT method<sup>41</sup>, relies on a variation of the Rapidly-exploring Random Tree (RRT)<sup>42</sup> and has been used in several studies for finding ligand unbinding pathways<sup>43,44</sup>. However, motion-planning-based methods still face two great challenges when dealing with protein-ligand interaction problems: high dimensionality and flexibility of the protein and ligand. Therefore, we propose in this article ART-RRT, a new method that combines the RRT method with the As-Rigid-As-Possible (ARAP) morphing technique from Computer Graphics to tackle these challenges. The ARAP technique has been successfully applied to find interpolation paths between two conformations of a molecular structure<sup>45</sup>. It has been shown to preserve well the rigidity of the original structures (in particular the bond lengths and bond angles) while allowing large motions during the interpolation. Here, the ARAP technique is used as a morphing and dimension-reduction method for the ligand, while an energy minimization is used to passively adapt the protein motions according to the ligand motions. As a result, the ART-RRT method efficiently produces low-energy pathways, which can then be refined further with path-optimization methods such as the Nudged Elastic Band<sup>46</sup> or String<sup>47</sup> methods for finding Minimum-Energy Paths (MEP). The paths produced by the ART-RRT method can also be used to start the Transition Path Sampling for computing free-energy differences<sup>48</sup> or reaction rate constants<sup>49</sup>.

# METHODOLOGY

The ART-RRT method combines the RRT method, which searches low-energy pathways, with the ARAP technique for dimension reduction, in order to find ligand unbinding pathways. We first review the RRT method and the ARAP technique, then introduce the ART-RRT method.

## Rapidly-Exploring Random Trees

RRT<sup>42</sup> is a stochastic-based motion planning method for exploring an arbitrary n-dimensional space. The method, originally developed for Robotics applications, has recently been applied to Structural Biology<sup>39,50</sup>. Essentially, RRT constructs a tree in a given state space. The tree is composed of nodes and edges. Each node represents a state while each edge connects a pair of nodes indicating a possible transition between these states. RRT has gained popularity thanks to the Voronoi-bias property which enables the tree growth preferably toward the unexplored regions of the space, thus avoiding the regions already visited.

The core of RRT is the extension mechanism relying on three steps (Figure 1). First, a state is randomly sampled within the search space. This state serves as the expansion direction for the current tree. Second, the nearest node in the current tree to the sampled state is obtained. Finally, the tree is extended from the nearest node toward the sampled state by generating new nodes between them (usually by linear interpolation). New nodes are accepted based on a certain condition. In Robotics, this condition usually rejects the states belonging to the collision (or obstacle) regions. However, different strategies can be used and the condition for the ART-RRT method will be discussed later.

Algorithm 1 describes the tree-extension process toward a target state  $q_t$ . The nearest node  $q_n$  is obtained at line 1 and a new node is generated by extending from  $q_n$  toward  $q_t$  at line 2. The *Extend*( $q_n, q_t$ ) function employs the linear interpolation to find  $q_{new}$ , i.e.,

$$q_{new} = \delta \frac{q_t - q_n}{\|q_t - q_n\|} + q_n \quad (1)$$

where  $\delta$  is the edge length or the extension step of RRT. If  $q_{new}$  is considered valid (line 3), it is added as a new node to the tree, and a new edge is created between  $q_{new}$  and the

nearest node (line 4 and 5). This extension scheme is repeated (line 6 and 7) as long as the new states are valid and the target state  $q_t$  is not reached (not indicated in the algorithm for clarity). The tree-extension process presented in Algorithm 1 is repeated until certain conditions are met such as when the tree has grown to a certain size.

---

**Algorithm 1:** ExtendToState function

---

**Input:** A Tree to extend  $\mathcal{T}$ , a target state  $q_t$ , an evaluation function **TestState**.

**Output:** A new branch added to  $T$ , with leaf node  $q_{new}$ .

1.  $q_n \leftarrow \text{NearestState}(\mathcal{T}, q_t);$
  2.  $q_{new} \leftarrow \text{Extend}(q_n, q_t);$
  3. **while**  $\text{TestState}(q_{new}) = \text{true}$  **do**
  4.      $\text{AddNode}(T, q_{new});$
  5.      $\text{AddEdge}(q_n, q_{new});$
  6.      $q_n \leftarrow q_{new};$
  7.      $q_{new} \leftarrow \text{Extend}(q_n, q_t);$
  8. **return**  $q_{new}$
- 

## As-Rigid-As-Possible Technique

The ARAP technique is commonly used in Computer Graphics to manipulate and animate meshes while preserving local rigidity such as edge lengths and angles. Since its introduction<sup>51</sup>, the development of the ARAP technique has received a lot of attention and led, in particular, to two types of three dimensional (3D) mesh manipulations: *interactive modeling* through actions on control points<sup>52–55</sup> and *shape interpolation*<sup>52,56–58</sup>. We have recently proposed the application of the latter for computing molecular interpolation paths<sup>45</sup>. We will now explain the notion of ARAP energy and its application to modeling as used in the ART-RRT method.

### ARAP Energy

The ARAP energy is a core concept in ARAP modeling and interpolation. Assume a mesh can be described by  $n$  vertices  $v_1, \dots, v_n$  connected by edges. In molecular structures, atoms

and bonds can be seen as vertices and edges, respectively. Let us consider an initial and a target state whose vertex positions are  $\mathbf{p}_i \in \mathbf{R}^3$  and  $\mathbf{p}'_i \in \mathbf{R}^3$ , respectively, for  $i \in [1, n]$ .

From this mesh, one can extract  $n$  sets  $\mathcal{N}_i$ , each of which consists of a central vertex  $v_i$  and all the vertices whose topological distances to this vertex is lower than a given value  $k$ . Here, we take  $k = 1$ , thus including only the one-ring neighbors which are the vertices directly connected to the central vertex  $v_i$ . Then, from each topological set, one can define a cell  $\mathcal{C}_i$  consisting of the vertex positions of the set  $\mathcal{N}_i$  for a given state.

Let us now consider  $\mathcal{C}_i$  and  $\mathcal{C}'_i$ , the cells centered on  $v_i$  in the initial and target states, respectively. Then, if the transformation  $\mathcal{C}_i \rightarrow \mathcal{C}'_i$  is rigid, there exists a rotation  $\mathbf{R}_i$  such that:

$$\mathbf{p}'_i - \mathbf{p}'_j = \mathbf{R}_i(\mathbf{p}_i - \mathbf{p}_j), \forall j \in \mathcal{N}_i \quad (2)$$

Hence, the approximate rigid transformation between the two cells can be defined as the rotation  $\mathbf{R}_i$  that minimizes the *cell deformation energy*  $E(\mathcal{C}_i, \mathcal{C}'_i)$ , also called *ARAP cell energy*:

$$E(\mathcal{C}_i, \mathcal{C}'_i) = \sum_{j \in \mathcal{N}_i} \omega_{ij} \|\mathbf{p}'_i - \mathbf{p}'_j - \mathbf{R}_i(\mathbf{p}_i - \mathbf{p}_j)\|^2 \quad (3)$$

where  $\omega_{ij}$  is the weight associated to edge  $e_{ij}$  connecting  $v_i$  with  $v_j$  in  $\mathcal{N}(i)$ . By default,  $\omega_{ij} = \omega_{ji} = 1$  for all edges. Figure 2 shows an example of a rotation  $\mathbf{R}_i$  which minimizes the ARAP cell energy.

When considering all the cells, the total ARAP energy  $E$  is defined as:

$$\begin{aligned} E &= \sum_i \omega_i E(\mathcal{C}_i, \mathcal{C}'_i) \\ &= \sum_i \omega_i \sum_{j \in \mathcal{N}_i} \omega_{ij} \|\mathbf{p}'_i - \mathbf{p}'_j - \mathbf{R}_i(\mathbf{p}_i - \mathbf{p}_j)\|^2 \end{aligned} \quad (4)$$

where  $\omega_i$  is the weight of set  $\mathcal{N}_i$  and its default value is  $\omega_i = 1$  for  $\forall i$ .

The minimization of the total ARAP energy is equivalent to minimizing the sum of all local deformations. As shown in Ref.<sup>45</sup>, the solution tends to preserve bond lengths and bond angles, and hence, lessen energy variations in molecular structures.



## ARAP Modeling

ARAP modeling is a useful method for 3D graphic design. A designer can obtain flexible shapes by constraining the motion of certain vertices on a mesh. The efficiency of the method allows real-time mesh editing. This mechanism has been proposed to manipulate molecular systems in the SAMSON software platform<sup>59</sup> as illustrated in Figure 3. It shows how a large deformation can be produced by displacing a very limited set of atoms, while preserving local rigidity.

The ARAP modeling method is described in Algorithm 2. The input consists of the initial vertex positions  $\mathbf{p}_i$ , the mesh edges, the set  $\mathcal{T}$  containing the constrained vertices, the constrained vertex positions  $\bar{\mathbf{p}}_i$  and the number of ARAP optimization steps  $m$ . First, the ARAP cells are generated for the initial state (line 2). The target vertex positions  $\mathbf{p}'_i$  are initialized by their corresponding values in the initial state except for the constrained vertices whose positions are initialized to  $\bar{\mathbf{p}}_i$  (line 3-6). Then,  $m$  steps of optimization are performed (line 7). At each step, the target cells are updated by the target vertex positions (line 9), which allows to compute the rotation that would preserve as much as possible the rigidity of the initial state (line 10). The new target vertex positions are computed (line 11) by minimizing the ARAP energy in Equation 4. The final target vertex positions are the output of the algorithm. The more iterations performed, the closer the result is to the initial shape, yet the constrained vertex positions are controlled.

More detailed descriptions of the `ComputeCell`, `ComputeRotation` and `ComputePositions` functions, also used in the ARAP interpolation process, can be found in Ref.<sup>45</sup>.

## ART-RRT

This article introduces the ART-RRT method that searches low-energy ligand unbinding pathways. It relies on the RRT method for efficiently exploring conformational spaces, and the ARAP modeling method for dimension reduction. Moreover, it uses constrained-optimization steps to update the protein shape according to the ligand motions. In ART-RRT, ligand atoms are labeled as either active or passive. The active-atom positions are randomly sampled by the RRT method, while the passive-atom positions are computed by

---

**Algorithm 2:** ARAP modeling

---

**Input:** The vertexes  $v_i$  with initial positions  $\mathbf{p}_i$ . The edges. The set  $\mathcal{T}$  of constrained vertices. The positions of constrained vertices  $\bar{\mathbf{p}}_i$  for  $v_i \in \mathcal{T}$ . The number of ARAP optimizations  $m$ .

**Output:** The new vertex positions  $\mathbf{p}'_i$ .

```
1. for  $i = 1, \dots, n$  do
2.    $\mathcal{C}_i \leftarrow \text{ComputeCell}(\mathbf{p}_i);$ 
3.   if  $v_i \notin \mathcal{T}$  then
4.      $\mathbf{p}'_i \leftarrow \mathbf{p}_i;$ 
5.   else
6.      $\mathbf{p}'_i \leftarrow \bar{\mathbf{p}}_i$ 
7.   for  $k = 1, \dots, m$  do
8.     for  $i = 1, \dots, n$  do
9.        $\mathcal{C}'_i \leftarrow \text{ComputeCell}(\mathbf{p}'_i);$ 
10.       $\mathbf{R}_i \leftarrow \text{ComputeRotation}(\mathcal{C}_i, \mathcal{C}'_i);$ 
11.       $\mathbf{p}'_1, \dots, \mathbf{p}'_n \leftarrow \text{ComputePositions}(\mathbf{R}_1, \dots, \mathbf{R}_n);$ 
```

---

the ARAP modeling method using the active-atom positions as constraints. Hence, thanks to ARAP modeling, the dimension of the search space is reduced to the number of the active-atom coordinates.

The general structure of the RRT method in Algorithm 1 remains valid for ART-RRT, but the **NearestState**, **Extend** and **TestState** functions need to be specialized, as described below.

### NearestState in ART-RRT

In standard RRT, state distances are evaluated in the nearest-node search and tree extension. The natural metric used is the RMSD taking into account all atoms of the system. In ART-RRT, because only active-atom positions are randomly sampled, this metric is therefore applied to only active atoms.

### Extend in ART-RRT

Standard RRT uses linear interpolation between the nearest node in the current tree to the target state to find  $q_{new}$ . In the ART-RRT extension, only the active-atom positions are linearly interpolated. The ARAP modeling method computes the passive-atom positions afterwards using the active-atom positions as constraints.

Although the states generated by the ARAP modeling method tend to have low energy, unrealistic states may still be produced, such as those where ligand atoms clash with protein atoms, since the protein conformation is not yet modified at this stage. Hence, an energy minimization is applied on the whole system (protein and ligand), relaxing it toward a local minimum. We use the FIRE method to perform the minimization because of its time and computational efficiency compared to common methods<sup>60</sup>. Also, the minimization is constrained to remain on the hyperplane orthogonal to the current expansion direction. This allows the states to relax while avoiding backtracking toward the parent conformations in the tree. This optimization process is illustrated in Figure 4.  $q_{new}^0$  is the state generated by the ARAP modeling method. The first minimization step gives  $q_{min}^0$  which is then projected on the hyperplane perpendicular to the direction made by  $q_n$  and  $q_{new}^0$ , giving the projected state  $q_{new}^1$ . By repeating the minimization and projection step, one obtains  $q_{min}^1$  and  $q_{new}^2$ . In ART-RRT, these operations are performed  $n_F$  times and lead to a final new state  $q_{new} = q_{new}^{n_F}$  which is then subjected to the **TestState** function described below.

### TestState in ART-RRT

Originally, RRT was developed for Robotics applications, where new states are considered valid if they do not lead a robot to collide with the environment. For biological applications, this test was adapted accordingly based on molecular energy, such as in the Transition-based RRT (T-RRT) method<sup>39,61</sup>. In T-RRT, a transition test based on the Metropolis criterion<sup>62</sup> is used with an adaptive mechanism, allowing the tree to favor low-energy regions while quickly escaping high-energy barriers. To take advantage of this feature, ART-RRT uses the same transition test for accepting/rejecting new states. Similar to Metropolis Monte Carlo methods<sup>63</sup>, a new state is accepted with a transition probability  $p_{ij}$

defined as:

$$p_{ij} = \begin{cases} \exp(-\frac{\Delta E_{ij}}{kT}) & \text{if } \Delta E_{ij} > 0, \\ 1 & \text{otherwise,} \end{cases} \quad (5)$$

where  $\Delta E_{ij} = E_j - E_i$  is the transition energy between the previous state  $q_i$  and the new state  $q_j$ .  $k$  is the Boltzmann constant.  $T$  is a temperature factor that does not necessarily carry any physical meaning and is considered as a parameter of the algorithm. The use of the temperature parameter is well known in a variety of molecular simulations such as Monte Carlo simulations<sup>63</sup> where the temperature is kept constant or Simulated Annealing simulations<sup>23</sup> where it is changed during the simulations. In our case, the temperature parameter is adaptively changed for controlling the difficulty of the transition test.

At the beginning,  $T$  is set to a low value to only allow the tree expansion on very easy positive slopes (in addition to flat and negative ones). Then, during the exploration, if the number of consecutive state rejections reaches a maximum value  $\mathcal{S}$  (for Severity), the temperature increases by a factor  $\lambda$  to ease the subsequent transition tests. In contrast, for each state acceptance, the temperature decreases by the same factor  $\lambda$ , thus making the subsequent tests more severe. Hence,  $T$  is automatically regulated along the exploration, balancing the search between the unexplored regions and low-energy regions.

## RESULTS and DISCUSSION

The ART-RRT method was implemented in C++ as a module of the SAMSON software platform<sup>59</sup>. Energy and forces were evaluated with the GROMOS96 43a1 force field in vacuum<sup>64,65</sup> using GROMACS<sup>66</sup> integrated in SAMSON.

For structure preparation, missing residues were modeled by MODELLER<sup>67</sup> integrated inside Chimera<sup>68</sup> and missing atoms by the swissPDB software<sup>69</sup>. Molecular topologies were generated by the *pdb2gmx* command in GROMACS for proteins and by the PRODRG server<sup>70</sup> for ligands. Before running ART-RRT, the systems were relaxed to their local minima using the FIRE method<sup>60</sup>.

The degrees of freedom are the Cartesian coordinates of the atoms. The parameters we used for ART-RRT are shown in Table 1. This choice was based on trials and errors on

several benchmarks and could probably be further improved, or tuned for specific scenarios. However, the optimization of these parameters is left for future investigations.

Table 1: Parameter setting used in ART-RRT.

Parameter description	Notation	Value
Number of ARAP iterations	$m$	20
Extension step in RRT tree	$\delta$	1 Å
Initial transition test temperature	$T$	0.001 K
Temperature factor in transition test	$\lambda$	2
Max number of failures in transition test	$\mathcal{S}$	1
FIRE integration time step	$t_F$	$1 \times 10^{-15}s$
FIRE number of steps	$n_F$	10

We applied the method to three cases of ligand unbinding pathways shown in Table 2. These benchmarks were chosen because their pathways had already been investigated by other well-known methods.

Table 2: Benchmarks used to evaluate ART-RRT

Id	Description	PDB	Chain Id	# ligand atoms	# protein atoms	Reference
I	Imatinib in protein kinase c-Kit	1T46	A	54	3178	<sup>15</sup>
II	Thiodigalactoside in lactose permease	1PV7	A	31	4292	<sup>11,41</sup>
III	Retinoic acid in human receptor	2LBD	A	22	2350	<sup>33</sup>

Since ART-RRT is based on a stochastic process, we ran the first two benchmarks 20 times each, and the third one 50 times where we observed a greater variety of pathways. For each benchmark, only two carbon atoms in the ligand were assigned as active, and their displacements were controlled by the RRT scheme. The other ligand atoms passively followed thanks to the ARAP modeling method. The protein atoms were passively moved thanks to the orthogonal minimization, except for one arbitrarily chosen atom whose position was always fixed, in order to avoid a global translation of the protein with the ligand as the

ligand escaped from the binding site. Each exploration was stopped as soon as the center of mass of a ligand was 40 Å away from its original position. Then, a solution path was extracted from the tree by linking the root node to the last accepted node. The sampling domain, *i.e.* the region where the search is performed for active atoms, was either a cubic volume centered on the ligand, or a rectangular volume restricted to the region of interest.

In benchmark I, we studied the influence of different sampling domains on ART-RRT results. In benchmark II, we analyzed the effect of post-processing ART-RRT paths with a path-optimization method.

## Unbinding of imatinib from the c-Kit protein kinase

This experiment involves imatinib, a type II kinase initially bound to the inactive form of the c-Kit protein kinase. A better understanding of the interaction mechanisms between kinase and its inhibitors is of major importance since they are involved in essential physiological processes<sup>71</sup>. This experiment is motivated by the study performed in Ref.<sup>15</sup>, which uses the SMD method to examine two candidate channels called ATP channel and allosteric-pocket (AP) channel.

Our goal is to show that ART-RRT is able to find these pathways at low computational cost, without the need for a reaction coordinate or an explicit bias. Since the setting we used largely differs from the SMD study (for example, we do not consider boundary conditions, explicit solvent nor constraints on alpha carbon atoms), we only expect to find the same types of pathways with potentially slight variations.

The initial complex and the active atoms on imatinib are shown in Figure 5. The figure also illustrates two sampling-domain setups that we used for this benchmark: a cubic region implying no direction-bias, and a rectangular region favoring solutions through the two candidate channels.

Figure 6 shows all the paths found by ART-RRT for the cubic and rectangular setups. In general, most paths belong to the candidate channels in both setups. However, other pathways were also found in the cubic setup, in addition to those along the candidate channels.

Table 3 presents a summary of the results by ART-RRT for the ATP and AP pathways as well as other pathways that do not follow these two roads. As seen from the table,

the average computational time to find a path is quite short compared with classical MD ( $416.4 \pm 47.6$  seconds for the cubic setup, and  $254.8 \pm 57.7$  for the rectangular setup). The rectangular setup took less time because the search did not wander in as large a space as in the cubic setup.

Table 3: Summary of results for Benchmarks I

	Cubic setup			Rectangular setup		
	# paths	Energy barrier (kJ.mol <sup>-1</sup> )	comput. time (s)	# paths	Energy barrier (kJ.mol <sup>-1</sup> )	comput. time (s)
ATP channel	5	$1600.3 \pm 214.0$	$403.9 \pm 27.4$	6	$1617.2 \pm 334.4$	$255.2 \pm 43.0$
AP channel	12	$1587.4 \pm 351.7$	$425.4 \pm 57.0$	14	$1427.9 \pm 294.0$	$254.6 \pm 62.9$
Other	3	$2995.4 \pm 1273.3$	$401.4 \pm 4.8$	0	-	-
Total	20	$1717.0 \pm 761.8$	$416.4 \pm 47.6$	20	$1484.7 \pm 318.7$	$254.8 \pm 57.7$

The table also shows that the paths in the “Other” category have significantly higher energy barriers compared to the rest, which explains why fewer paths are found for this category. In addition, the AP paths have generally smaller energy barriers than the ATP paths for both setups.

Figure 7 (for the rectangular setup) and Figure 8 (for the cubic setup) help us understand why the AP pathway tends to have smaller energy barrier than the ATP pathway. Each plot in these figures shows the mean and standard deviation of the maximum displacement of the alpha carbons from the initial bound state, for either the ATP or AP paths. According to previous studies<sup>15,72,73</sup>, there are four regions which move significantly during the unbinding process of imatinib from the c-Kit protein (see Figure 9): the JMR (GLY-1 to ASP-15), the  $\beta$ -sheet (LEU-25 to LEU-61), the helix  $\alpha$ C (HIS-66 to GLY-84) and the A-loop (CYS-187 to LEU-209). Note that our residue numbering is different from the one of the original pdb file, due to the reconstruction of the missing residues. As shown in both figures, these regions are also found to be the most mobile in the ART-RRT results. In the rectangular setup, the three most mobile regions are the  $\beta$ -sheet, helix  $\alpha$ C, and A-loop for the ATP paths (Figure 7a) while all of the four mentioned regions are found to be most mobile in the AP paths (Figure 7b). The same behavior is observed for the cubic setup (Figure 8) except that more mobile regions show up and the RMSD values are also smaller in general compared with the results in the rectangular setup. This is because the cubic setup allows

for more exploring directions, and hence, the ligand can push less on the residues of the important regions while pushing more on other residues on the channel lining. In both figures, larger displacements of the alpha carbon atoms in the ATP paths than in the AP paths are observed. For example, the maximum mean value in Figure 7 for the ATP paths is greater than 3 Å, and for the AP paths is less than 2 Å. This implies that several residues must be displaced to a greater extent to facilitate the ligand passage along the ATP channel than the AP channel.

Let us now examine in detail the candidate pathways. For this purpose, two representative paths are picked from the rectangular setup, i.e. paths which give the similar patterns to those of the mean values in Figure 7. One of them represents the ATP path and the other represents the AP path. Figure 10 shows the average displacement of the alpha carbons of the most mobile residues (only the alpha carbons which have the maximum displacement more than 0.5 Å are considered) and the Van der Waals (VdW) energy along the paths. The curves in the left plot rise from 0 Å to a maximum value, indicating the opening of the channels, then fall to stable levels, indicating the closing of the channels after the ligand escape. Our results agree with the SMD study that the passage along the ATP channel leads to more displacement for the residues involved. Our interpretation is that the passage along the ATP channel, therefore, requires more energy (see Table 3) to push the residues obstructing the channel. The only difference in the left plot of Figure 10 with that in the SMD study is that our measurement is lower in value, probably due to the lack of water molecules in our experiment. Note that the water molecules can keep the channels open wider by filling the ligand place after its escape. The right plot in Figure 10 shows the VdW energy of these representative paths. The VdW energy barriers of our curves are about  $0.03 \times 10^4$  kcal/mol, which is close to that found by the SMD study ( $0.02 \times 10^4$  kcal/mol). Interestingly, similar to the SMD study, the VdW energy barrier of the ATP path also occurs before that of the AP path.

The results from our experiment support the hypothesis that the AP channel is preferred for the ligand unbinding. The same conclusion is found in a recent study of the same protein family using the umbrella sampling simulation<sup>74</sup>. However, whether the ATP or AP channels are preferred for the ligand unbinding is still debatable since a different conclusion is reached



by the SMD study<sup>15</sup>. The differences between our conclusion and that from the SMD study may come from the experiment setup. In particular, our experiment is done in vacuum, while the SMD study is done with explicit solvent. Secondly, the chosen direction of the pulling force in the SMD method may not be ideal for the escape of the ligand while our method is sampling-based and favors the passage along low-energy regions. In any case, the main purpose of this experiment is to show the capability of our method to efficiently find the main pathways found by other methods from the literature.

## Unbinding of Thiodigalactosid from Lactose permease

We used ART-RRT to simulate the unbinding of Thiodigalactosid from the Lactose permease, a twelve-alpha-helical membrane transport protein<sup>75</sup>. The goal of this experiment is to compare ART-RRT with the ML-RRT method proposed by Ref.<sup>41</sup>. The ML-RRT method, which also relies on the RRT exploration scheme, represents both the ligand and the protein in internal coordinate system, *i.e.* a set of dihedral angles. For this benchmark, ML-RRT allows full flexibility of the ligand, whereas only certain parts of the protein are flexible. ML-RRT also divides the whole system in passive and active parts, where the active parts are controlled by the RRT scheme and the passive parts are displaced as soon as steric collisions are detected with the active parts.

The sampling region and active ligand atoms for ART-RRT are shown in Figure 11. This region is limited to the upper part of the protein since we want to study the ligand unbinding toward the periplasmic side of the protein only.

For this benchmark, ART-RRT takes  $136.5 \pm 21.1$  seconds to find each path versus 1 hour for the ML-RRT method. In average, this corresponds to a speed-up of more than 26 times compared with the ML-RRT method. The computational time is also much smaller than what is required for classical MD simulations.

For this study, we also investigate the effect of a path-optimization method to locally improve the paths obtained with the ART-RRT method. Therefore, we apply the Nudged Elastic Band (NEB) method<sup>46,76</sup> to optimize the ART-RRT paths. The NEB method is implemented as a parallel module in the SAMSON software platform<sup>59</sup>. To ensure the stability of the NEB results, we only keep a limited number of points along the ART-

RRT paths<sup>77</sup>. Hence, each ART-RRT path is cut down to about 28-38 conformations per path before the NEB method is applied. The total computational time to obtain a path with ART-RRT that are later optimized with the NEB method is  $181.0 \pm 30.2$  seconds. This post-treatment operation is, hence, computationally cheap (about 44.5 seconds more is spent for each path) and leads to much lower energy barriers (2 to 16 times), as shown in Figure 12. However, this optimization only adjusts a given path *locally*, while its nature remains unchanged (see Figure 13).

To compare with the ML-RRT method, we recorded the contacts that the ligand makes with the protein along its unbinding pathway. As defined in Ref.<sup>41</sup>, a contact is recorded when the distance between a protein atom and a ligand atom is lower than the sum of their Van der Waals radii plus 1 Å.

The top part of Figure 14 shows the probability of contact between the ligand and the protein during unbinding, for the residues reported in Ref.<sup>41</sup>, along three segments of the pathway: 0-10 Å, 10-20 Å, and after 20 Å. Precisely, each box indicates the percentage of paths in which a contact is present between the ligand and a particular residue for a given path segment. We observe that all the contacts reported by the ML-RRT method are also found with the ART-RRT method. Moreover, the contact patterns are similar: for example, residues GLU-269 to ASP-237, PHE-27 to ASN-245 and THR-45 to HIS-35 (from left to right in the figure) appear at the beginning, the middle and the end of the unbinding path, respectively. Interestingly, this list of residues are also reported to have hydrogen bonding and hydrophobic interactions with the ligand in another study using the SMD method<sup>11</sup>.

The bottom part of Figure 14 shows more contacts found by ART-RRT which appear at least 30 % of the paths. In Ref.<sup>11</sup>, the residues GLU-269, HIS-322, ARG-144, ARG-302, GLU-325 and GLU-126 are deemed essential for the lactose transport. ART-RRT detected the interaction of the ligand with 5 out of these 6 residues (GLU-269, HIS-322, ARG-114, ARG-302, GLU-325) whereas only GLU-269 and HIS-322 are reported in the ML-RRT results. GLU-325 is not shown in the figure because its presence is less than 30 % of the paths while GLU-126 was not considered in our experiment setup because it does not belong to the passage toward the periplasmic side of the protein.

We also analyzed the effect of the NEB method on the contact pattern (see Figure 15).

The comparison between this figure and Figure 14 shows that the contact pattern is essentially preserved although the NEB method reduces the energy barriers of the paths significantly (see Figure 12). Five contacts (GLN-359, MET-323, MET-362, PHE-246 and TYR-373) become less present (less than 30 % of the path), and hence, do not show up in this figure. Only one contact (with TYR-26) is no longer present in the paths after the optimization.

Hence, the protein-ligand contact analysis shows that ART-RRT can give results comparable to those obtained with either the ML-RRT or the SMD method in the setup proposed in<sup>11</sup>, but in a much shorter computational time. Moreover, the contact analysis and Figure 13 shows that although the NEB method remarkably reduces the energy barriers of the ART-RRT paths, it does not significantly change the path natures.

Figure 16 shows the maximum displacements of the alpha carbons along the unbinding pathway for the paths before and after NEB optimization. As one can see, the most mobile residues are PHE-20 to PHE-55, ILE-230 to LEU-271, and MET-365 to LEU-385. The displacements of the residues from PHE-20 to PHE-55 and ILE-230 to LEU-271 are not surprising because these residues lie on the channel lining. The displacements of the residues from MET-365 to LEU-385 are induced by the motion of the residues ILE-230 to LEU-271. The figure also shows the effect of the NEB method that may reduce the RMSD values of some of the most mobile residues while slightly increasing the mobility for few other residues.

To observe how the narrowest constriction of the channel (made by the gap between residues ILE-40 and ASN-245) reacts to the ligand-unbinding event, we measured the maximum alpha-carbon distances between these two residues along the unbinding paths. We found that its maximum opening is  $12.1 \pm 0.4 \text{ \AA}$  and  $12.3 \pm 0.4 \text{ \AA}$  for the paths before and after optimization, respectively. This result is slightly smaller than  $15 \text{ \AA}$  that is the distance found by the ML-RRT method and reported by another experimental study<sup>78</sup> as necessary for lactose transport. The difference may be due to two reasons. First, our method lets the protein react according to the potential forces, and hence, large motions of the protein are not addressed. Second, water molecules which can widen the channel by taking the ligand place during the unbinding process are not modeled in our study. Despite this quantitative discrepancy, the paths found by our method show similar characteristics with those produced

by the ML-RRT method and the SMD method<sup>11</sup>.

## Unbinding of retinoic acid hormone from its receptor

Nuclear hormone receptors are involved in many cellular processes such as reproduction, transcription, etc. and hence, subjected to many researches<sup>16,33</sup>. Here, we study the unbinding pathways of retinoic acid from its receptor as in Ref.<sup>33</sup>, where the SMD method was used. The bound state of the protein-ligand complex is modeled from PDB entry 2LBD. Figure 17 shows the two active atoms on the ligand and their cubic sampling region. As noted above, since a large variety of pathways were found in this benchmark, ART-RRT was run 50 times in order to produce averages.

In Ref.<sup>33</sup>, three pathways (I, II, III) were chosen for the SMD simulations based on the bound structure of 2LBD. Pathway I occurs through the space between H11, H12 and the loop made by them. Pathway II is through the space beneath H11 and H12. In our result, we also include the space between H3 and the loop H11-H12 in pathway II. Pathway III is a tunnel that can be seen by looking at the molecular surface between H3 and the loop H1-H3. The ART-RRT method was capable to find all of these pathways (see the left picture of Figure 18).

Table 4 shows the number of paths found for each pathway as well as the average energy barrier and computational time. The ART-RRT method spent about  $399 \pm 59.1$  seconds to find each path.

In addition to the mentioned pathways, ART-RRT also found pathways IV, V, VI and Other (pathways which do not belong to the other six categories). Pathway IV is through the space between H11 and the N-terminal of H7. Pathway V is between H6, H7 and the  $\beta$ -sheet between H5 and H6. Pathway VI is through the space between the C-terminal of H1 and the  $\beta$ -sheet between H5 and H6. All the ART-RRT paths along these pathways can be seen in the right picture of Figure 18. Interestingly, pathways IV, V and VI are also reported by another study that employs the RAMD method to find ligand unbinding pathways for another nuclear hormone receptor<sup>16</sup>. This shows the efficacy of the ART-RRT method in finding a large diversity of candidate pathways in just a few minutes for each path.

Table 4: Summary of results for Benchmarks III

	#	Energy barrier	comput.
	paths	(kJ.mol <sup>-1</sup> )	time (s)
Path I	10	943.8 $\pm$ 313.3	411.8 $\pm$ 57.7
Path II	6	933.8 $\pm$ 338.0	364.6 $\pm$ 19.6
Path III	13	907.2 $\pm$ 207.0	390.1 $\pm$ 61.1
Path IV	14	895.1 $\pm$ 208.4	383.3 $\pm$ 59
Path V	4	1405.6 $\pm$ 182.9	471.4 $\pm$ 50.9
Path VI	1	1201.7	474.9
Other	2	1466.0 $\pm$ 216.5	422.5 $\pm$ 8.2
Total	50	982.4 $\pm$ 300.0	399.0 $\pm$ 59.1

## CONCLUSIONS

ART-RRT is a new method for efficiently computing approximate ligand unbinding pathways. The method is based on the RRT scheme for exploring conformational spaces, and the ARAP technique for dimension reduction. Moreover, ART-RRT employs energy evaluation and local optimization to produce low-energy paths. The method was applied on three benchmarks where the data were available for comparison. Our experiments showed that the ART-RRT method is fast and able to find a diversity of low-energy pathways. The method can also be easily tuned to focus the search in specific spatial regions. Overall, the results are in good agreement compared to those found by the state-of-the-art approaches.

The paths found by ART-RRT could be further refined to compute minimum-energy paths, or used by other advanced methods such as the transition path sampling to generate a path ensemble and estimate free-energy differences or reaction rate constants.

Despite the preliminary success for the presented benchmarks, the current method still have several drawbacks such as the inability to address large protein motions and handle explicit solvents. Therefore, we are considering several directions to improve and complement the current method. Firstly, the investigation on how the method parameters, as well as the location and number of active atoms, affect the results is necessary. At the moment, only

two active atoms are selected and located at the extremities of the ligand. Secondly, the placement of active atoms on the proteins is an interesting strategy to sample large protein motions during the unbinding process. Thirdly, the adaptation of the current method for solvated systems will be beneficial for many users. Fourthly, we would like to extend the method for more complex problems such as conformational changes of a protein and protein-protein interactions. Finally, the method can be extended to use multiple exploration trees for finding the pathways between two or more given states.

## **ACKNOWLEDGMENTS**

We would like to gratefully acknowledge funding from the European Research Council through the ERC Starting Grant No. 307629.

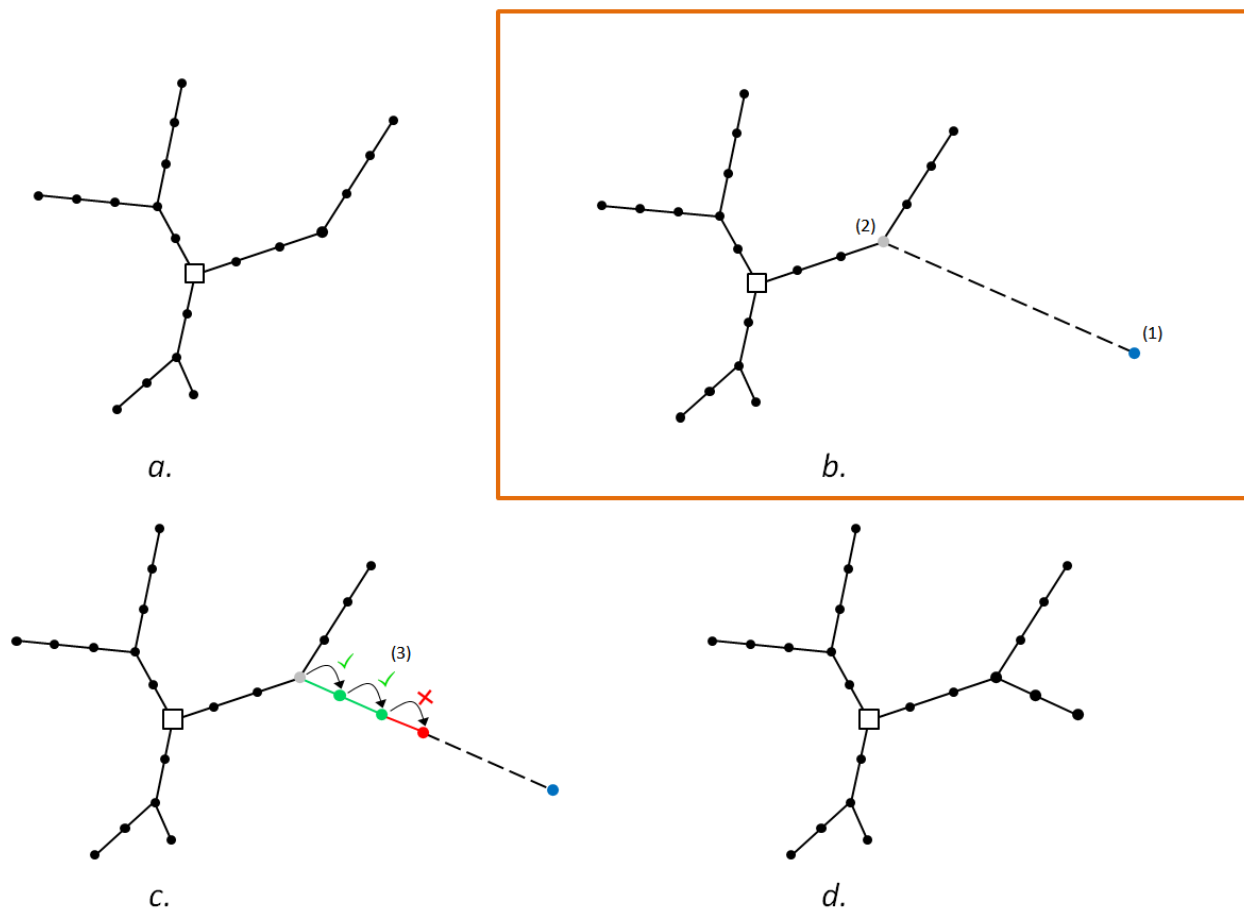


Figure 1: An extension step in RRT. *a.* The tree at a given stage of the search. The white square represents the root node. *b.* A conformation (1) is randomly sampled in a conformational space represented by the orange box (this conformation can be unrealistic, its role being only to provide a direction) and the nearest node from this conformation (2) is found in the tree. Note that the space bounds are typically set to be much larger compared with the volume covered by the tree. *c.* Several extension steps are attempted (3) from the nearest node toward the sampled state based on some acceptance criterion. The expansion stops as soon as a new state is rejected. *d.* The resulting tree after the extension step. This extension scheme is known to explore preferably new regions of the space.

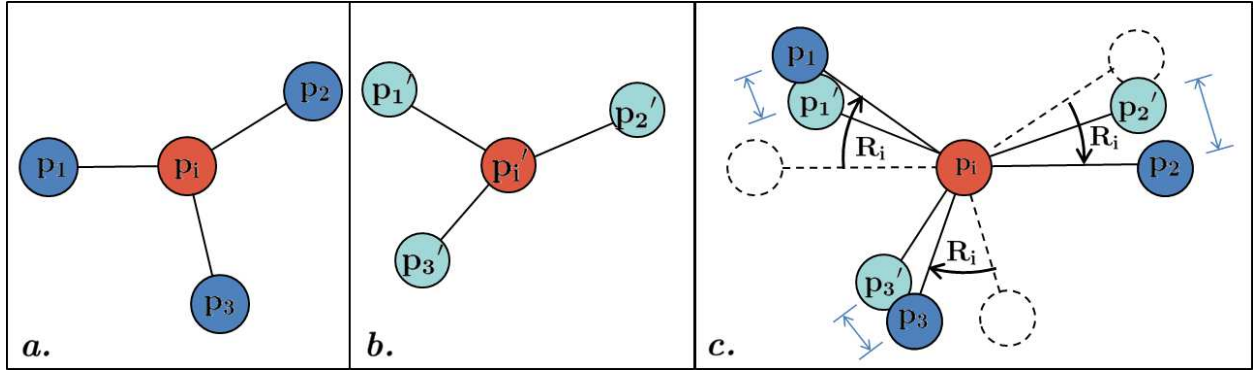


Figure 2: The local ARAP rotation that maximizes rigidity. *a.* The initial cell. *b.* The target cell. *c.* The optimal rotation  $\mathbf{R}_i$  minimizes the sum of the rotated  $\|\mathbf{p}_j - \mathbf{p}'_j\|^2$  distances (arrow bars), after aligning the cells such that  $\mathbf{p}_i = \mathbf{p}'_i$ .



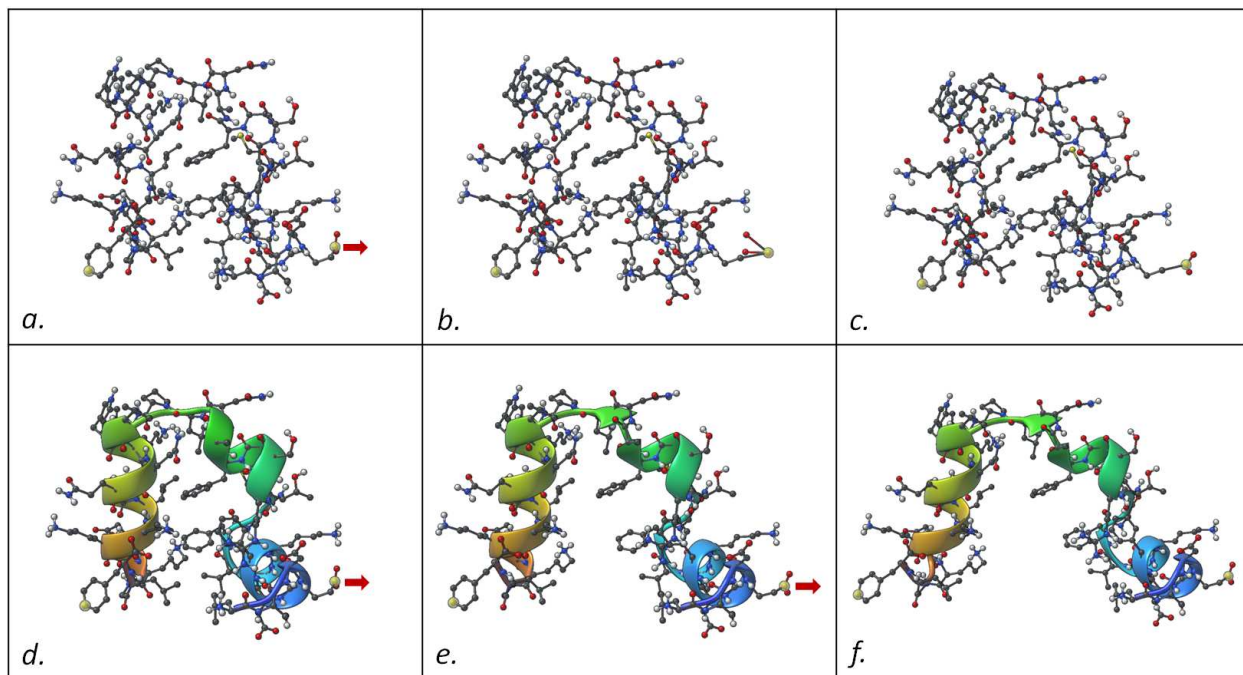


Figure 3: ARAP modeling applied on a molecular structure (PDB id: 1YRF) as simulated in the SAMSON software platform. *a.* The two atoms in yellow at the bottom left and right are constrained. *b.* The right constrained atom is displaced and the left one is fixed. The picture shows the resulting shape in the case where the ARAP modeling is not applied. *c.* Thanks to ARAP modeling, the displacement leads to the entire structure modified while the structural rigidity is preserved as much as possible and the position of the left constrained atom stays fixed. *d.*, *e.* and *f.* Large deformations of the system by successive motions of the right constrained atom. Note how the secondary structure is largely preserved.

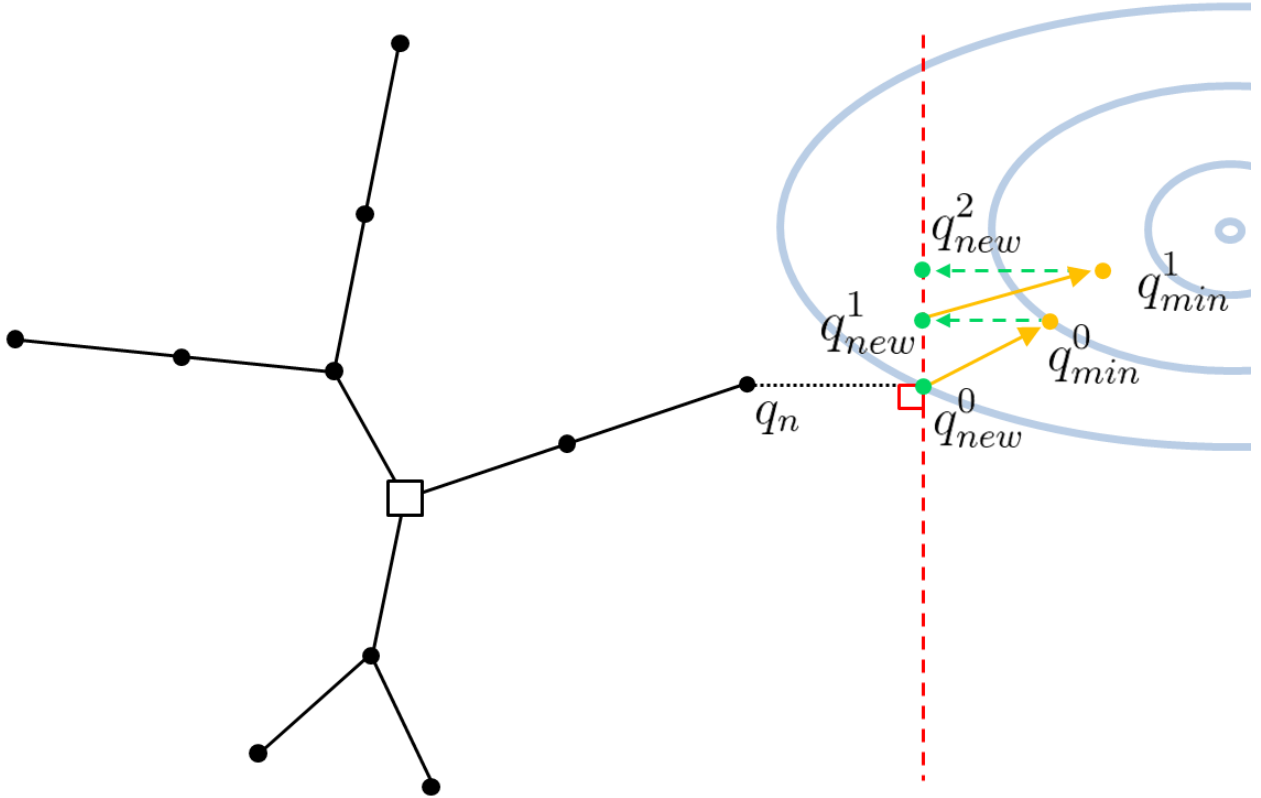


Figure 4: FIRE orthogonal update with 2 iterations.  $q_{new}^0$  is the state generated by ARAP modeling.  $q_{min}^0$  is the first minimized state using the FIRE method.  $q_{new}^1$  is the projected state of  $q_{min}^0$  onto the plane perpendicular to the direction made by  $q_n$  and  $q_{new}^0$ . This plane is represented by the red dotted line.  $q_{min}^1$  and  $q_{new}^2$  are the minimized and projected states, respectively, for the second step.

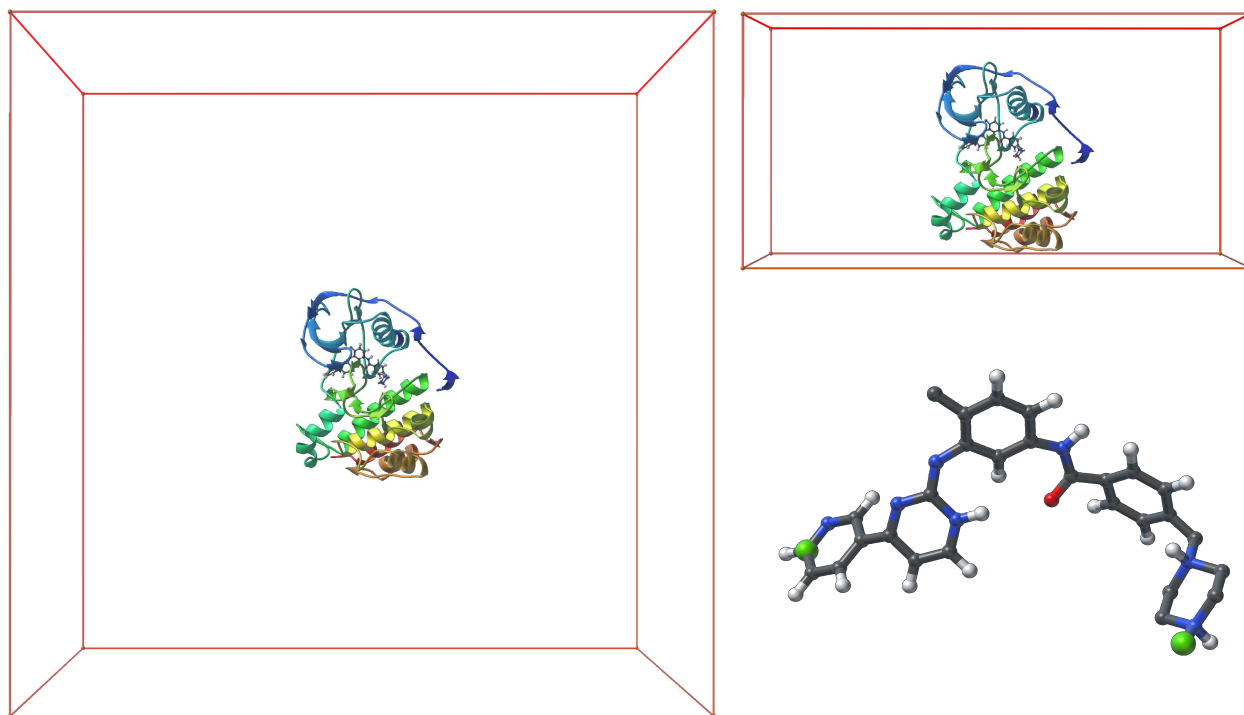


Figure 5: Left (the cubic setup): the protein-ligand complex for Benchmark I at its initial state inside a cubic box (centered on the ligand) defining the sampling domain for the active atoms. Top-right (the rectangular setup): the complex inside the rectangular sampling domain centered on the ligand. Bottom-right: A closer view on the ligand (imatinib) where two carbon atoms (in green) are set as active.

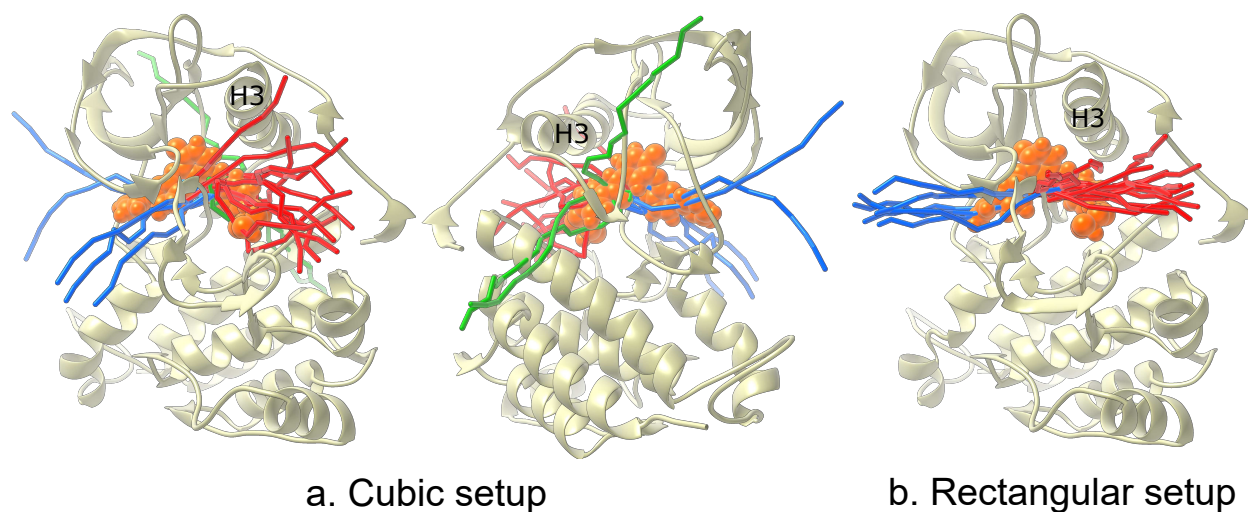


Figure 6: The ART-RRT paths for both setups in Benchmark 1. Views from both sides of the protein are shown for the cubic setup. The protein is represented by the ribbons and the ligand in orange balls. The ATP and AP paths are represented by the blue and red sticks, respectively. The other paths are represented by the green sticks. Each stick traces the center of mass of the ligand.

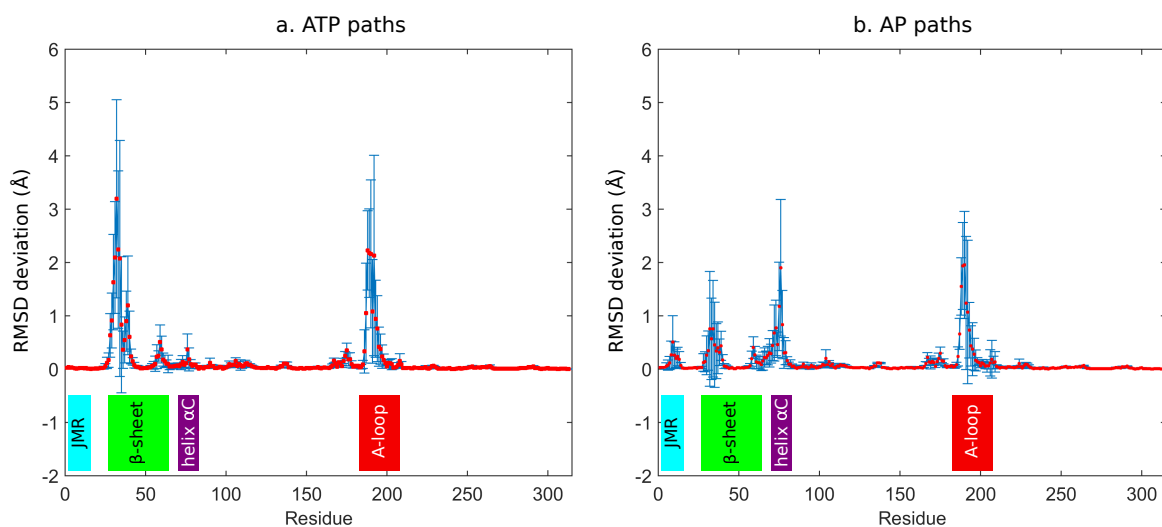


Figure 7: Benchmark 1 with the rectangular setup: The mean (red dots) and standard deviation (vertical blue bars) values of the maximum RMSD deviations from the initial bound state of alpha carbons for a) the ATP paths and b) the AP paths. The residues of 4 important regions (JMR,  $\beta$ -sheet, helix  $\alpha$ C and A-loop) are spanned by the colored boxes. The AP paths involve motions in all these 4 regions while the ATP paths involve motions in all of them except the JMR.

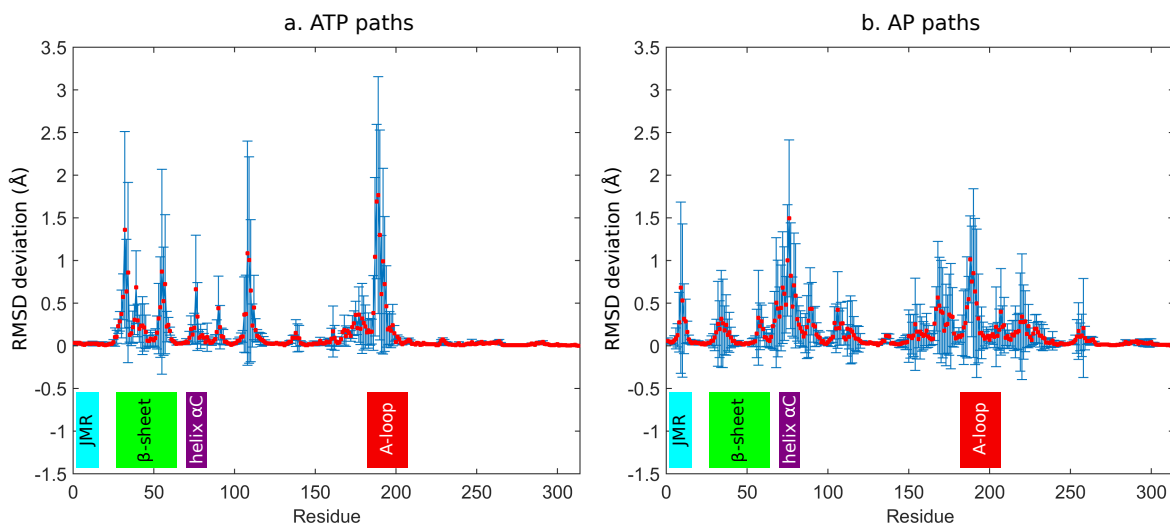


Figure 8: Benchmark 1 with the cubic setup: The mean (red dots) and standard deviation (vertical blue bars) values of the maximum RMSD deviations from the initial bound state of alpha carbons for a) the ATP paths and b) the AP paths. The residues of 4 important regions (JMR,  $\beta$ -sheet, helix  $\alpha$ C and A-loop) are spanned by the colored boxes. Similar to the rectangular setup, the AP paths involve motions in all of the 4 regions while the ATP paths involve motions in all of them except the JMR. Besides these important regions, the cubic setup leads to significant motions in other regions.

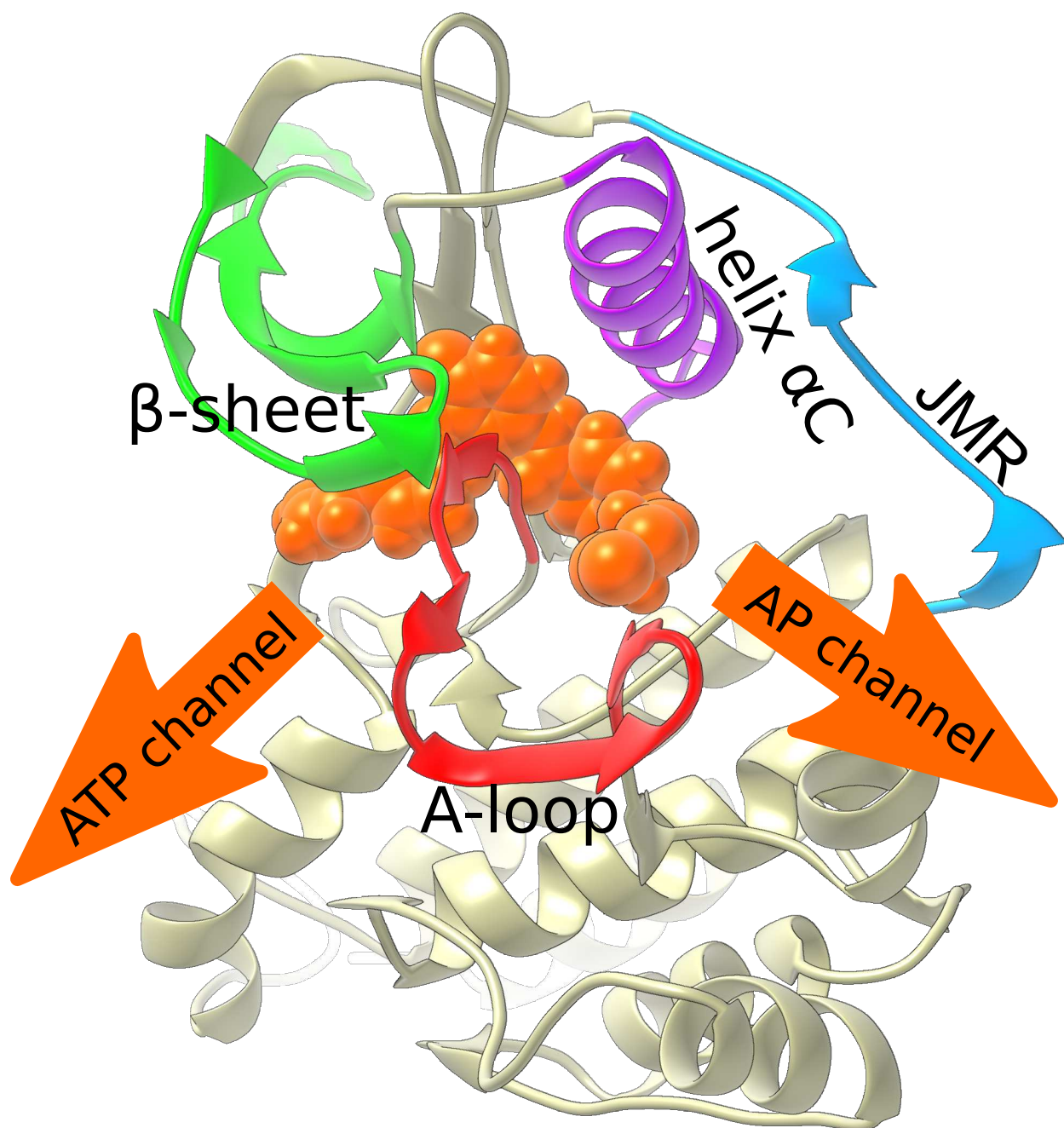


Figure 9: Location of the JMR (blue),  $\beta$ -sheet (green), helix  $\alpha C$  (purple) and A-loop (red). The ligand is represented by orange balls.

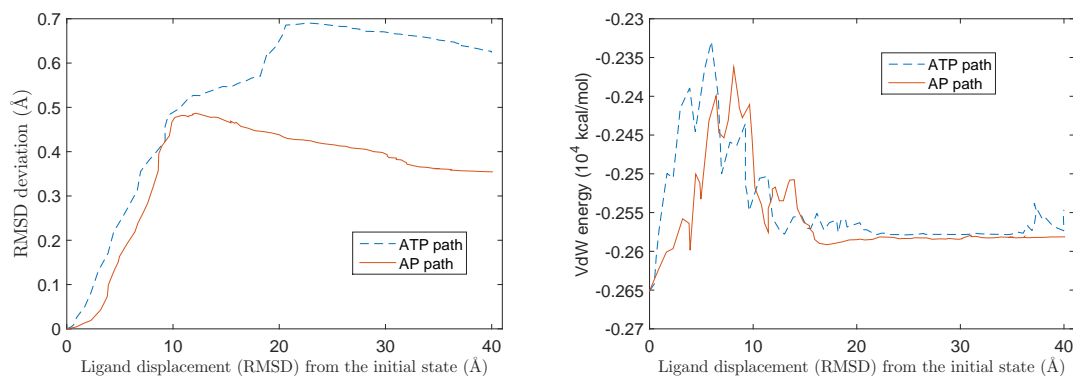


Figure 10: Results for the representative ATP path and AP path of the Rectangular setup. The left plot represents the average RMSD deviation of the alpha carbons composing the ATP and AP channels from the initial bound states along the dissociation pathways. One can see higher RMSD values for the ATP path, implying more motion of the residues to give way for the ligand along this channel. The right plot shows the Van der Waals energy along the dissociation pathways. Similarly to the result from Ref.<sup>15</sup>, it appears that the energy barrier of the ATP path occurs before that of the AP path.



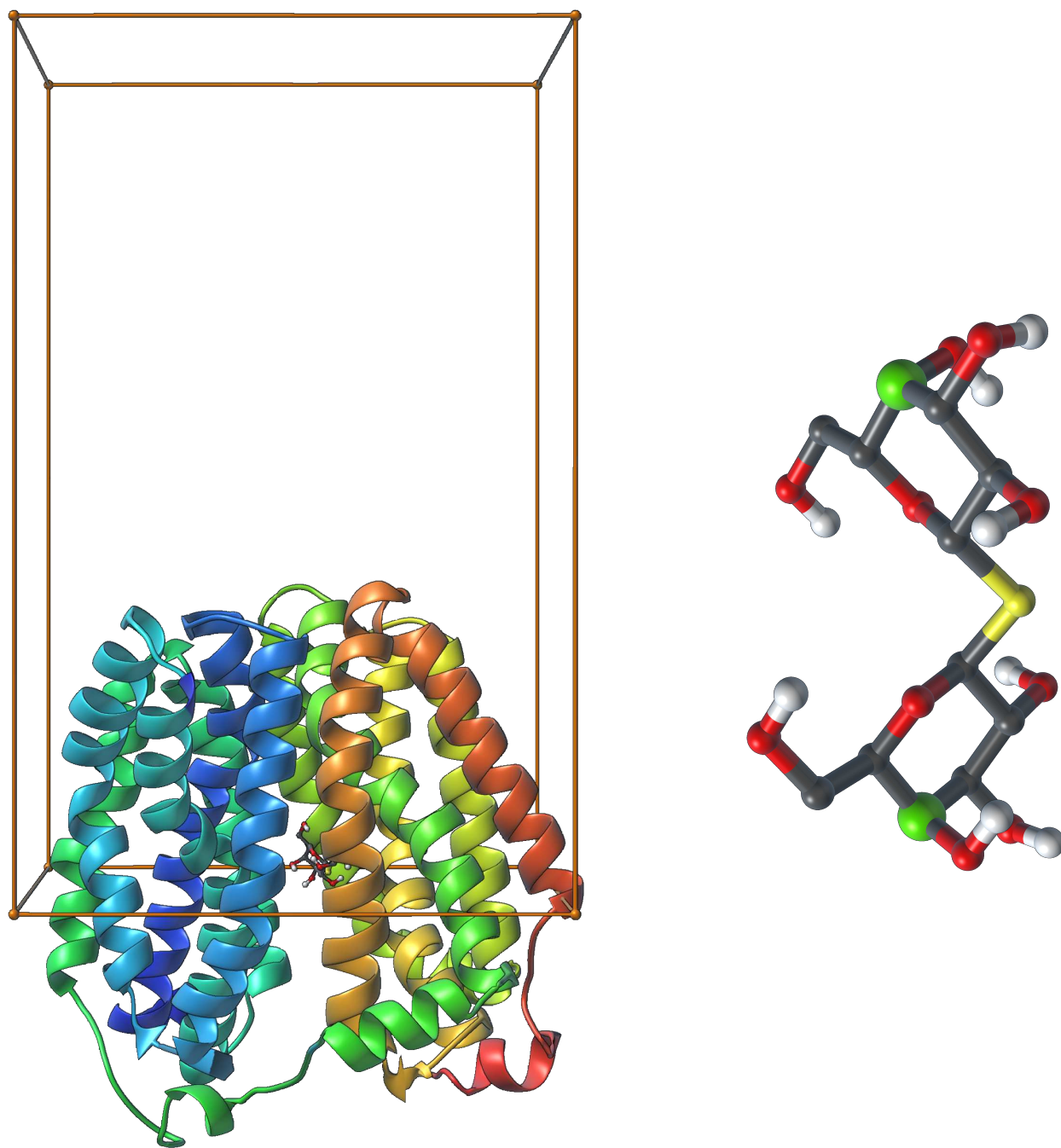


Figure 11: Left: The ligand-protein complex for Benchmark II (in ribbon) with the sampling box for active atoms. This box is biased toward the periplasmic side of the protein. Right: a closer view on the ligand with the two active carbon atoms in green.

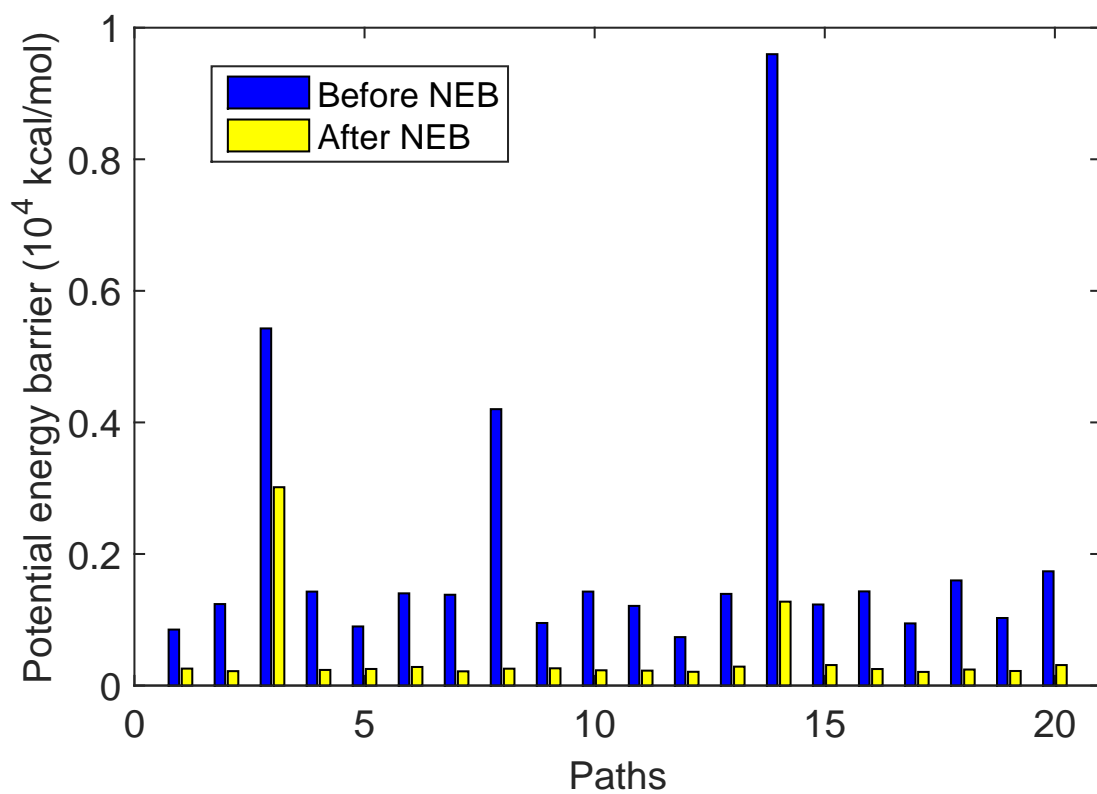


Figure 12: Potential energy barriers for the paths directly obtained with ART-RRT and those post-processed with the NEB method. This optimization step greatly reduces the energy barriers (from 2 to 16 times).

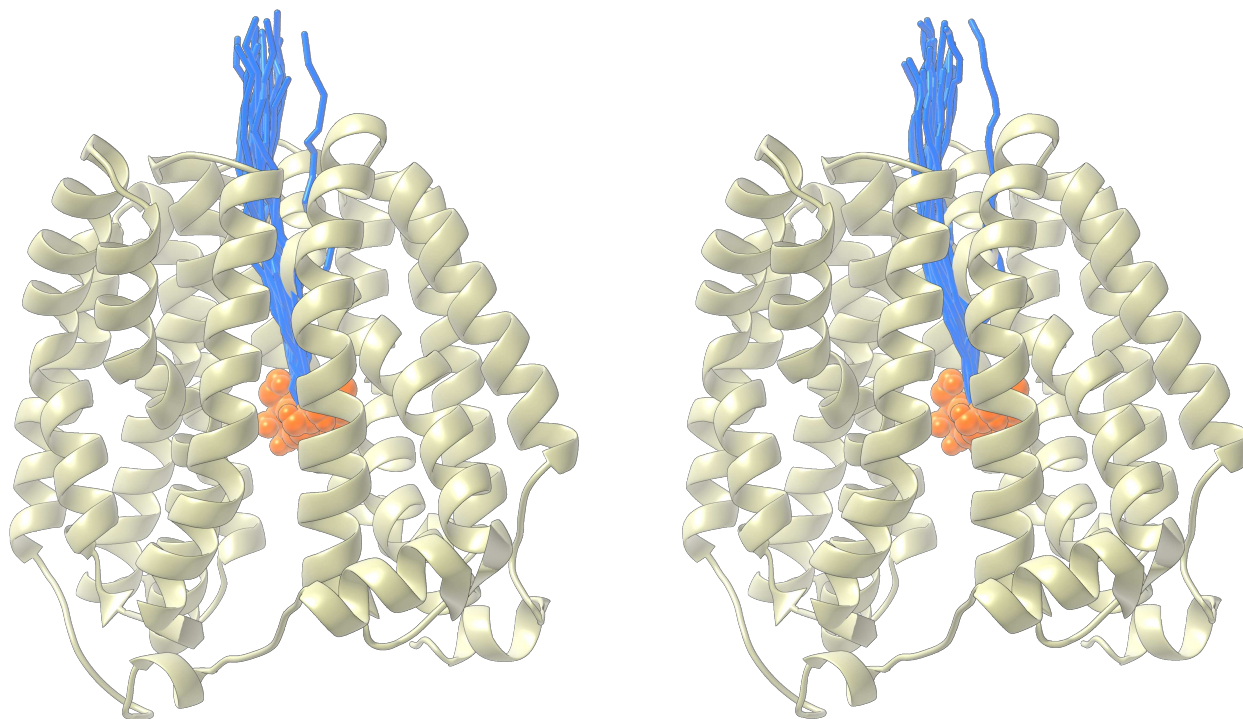


Figure 13: All the paths found for Benchmark II (in blue sticks). Left: paths obtained by ART-RRT. Right: optimized paths obtained by ART-RRT + NEB. The protein and ligand are represented by ribbons and orange balls, respectively. The NEB optimization only adjust the paths locally, while the path natures remain unchanged.

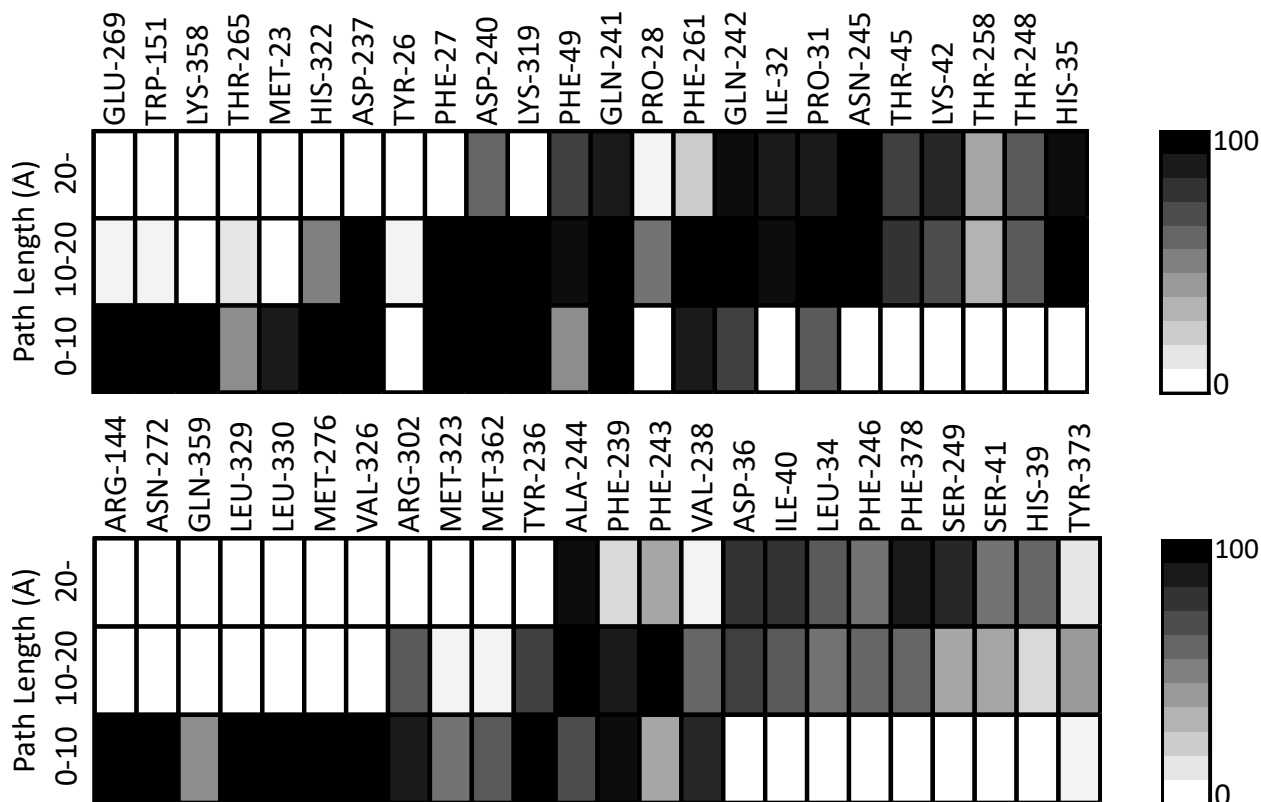


Figure 14: Contacts made by the protein residues and the ligand along 3 segments of the unbinding paths (0-10 Å, 10-20 Å, and after 20 Å) found by ART-RRT. The grey scale shows the percentage of times a specific contact is present for that particular path segment over all pathways. On the top is the contact result with the residues also found by ML-RRT. On the bottom is the contact result with other residues found by ART-RRT. Only residues which have contact at least 30 % of the paths are shown.

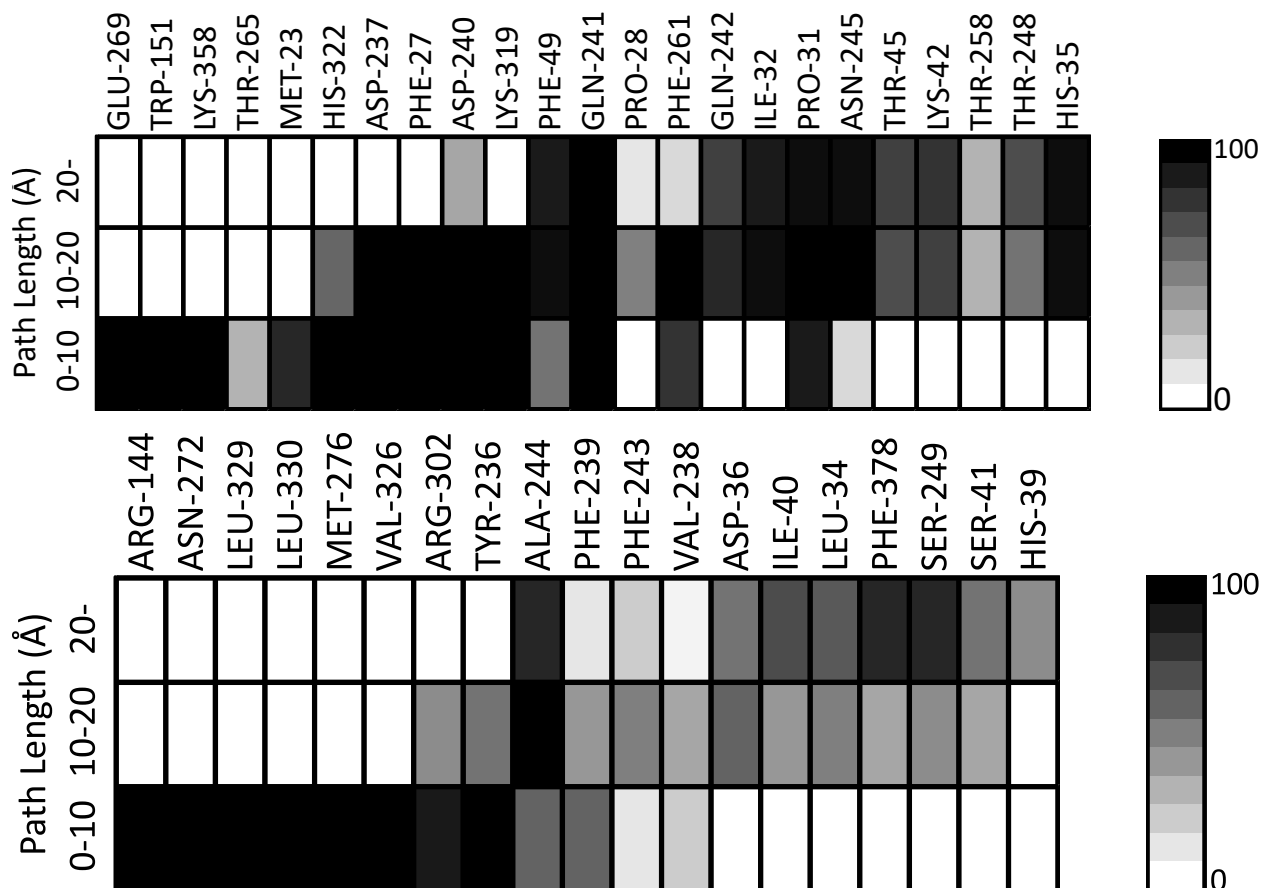


Figure 15: Contacts between the protein residues and the ligand along 3 segments of the optimized paths (0-10 Å, 10-20 Å and after 20 Å) after NEB. The grey scale shows the percentage of times a specific contact is present for that particular path segment over all pathways. On the top is the contact result with the residues also found by ML-RRT (TYR-26 is no longer in this list). On the bottom is the contact result with other residues found by ART-RRT. Only residues which have contact at least 30 % of the paths are shown. GLN-359, MET-323, MET-362, PHE-246 and TYR-373 are no longer in this list.

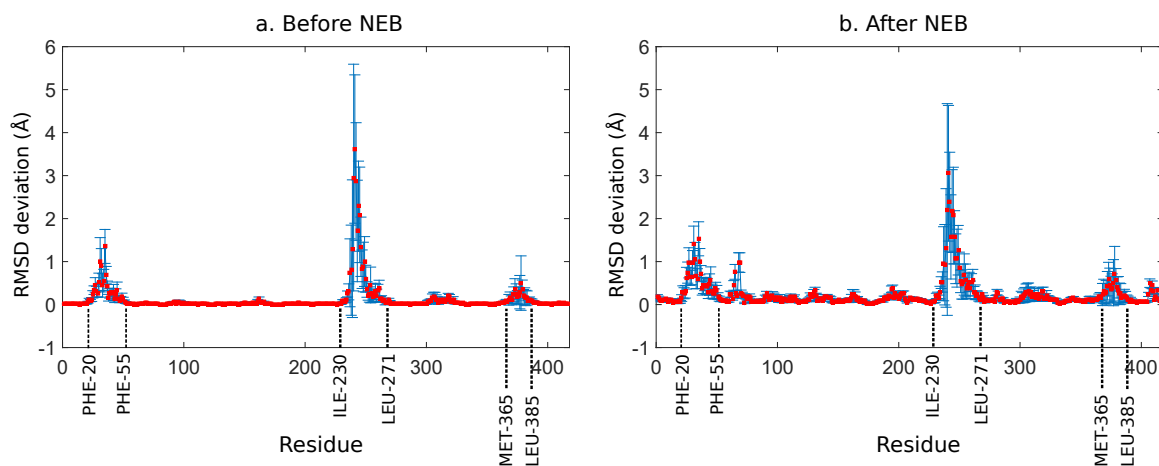


Figure 16: The mean (red dots) and standard deviation (blue vertical bars) values for the maximum displacement of the alpha carbons from the initial binding state before and after optimization. The residues PHE-20 to PHE-55, ILE-230 to LEU-271, and MET-365 to LEU-385 are moved the most by the ligand during the unbinding. The paths after the optimization lead to lower RMSD values for the most mobile region, but more regions are subjected to some small displacements.

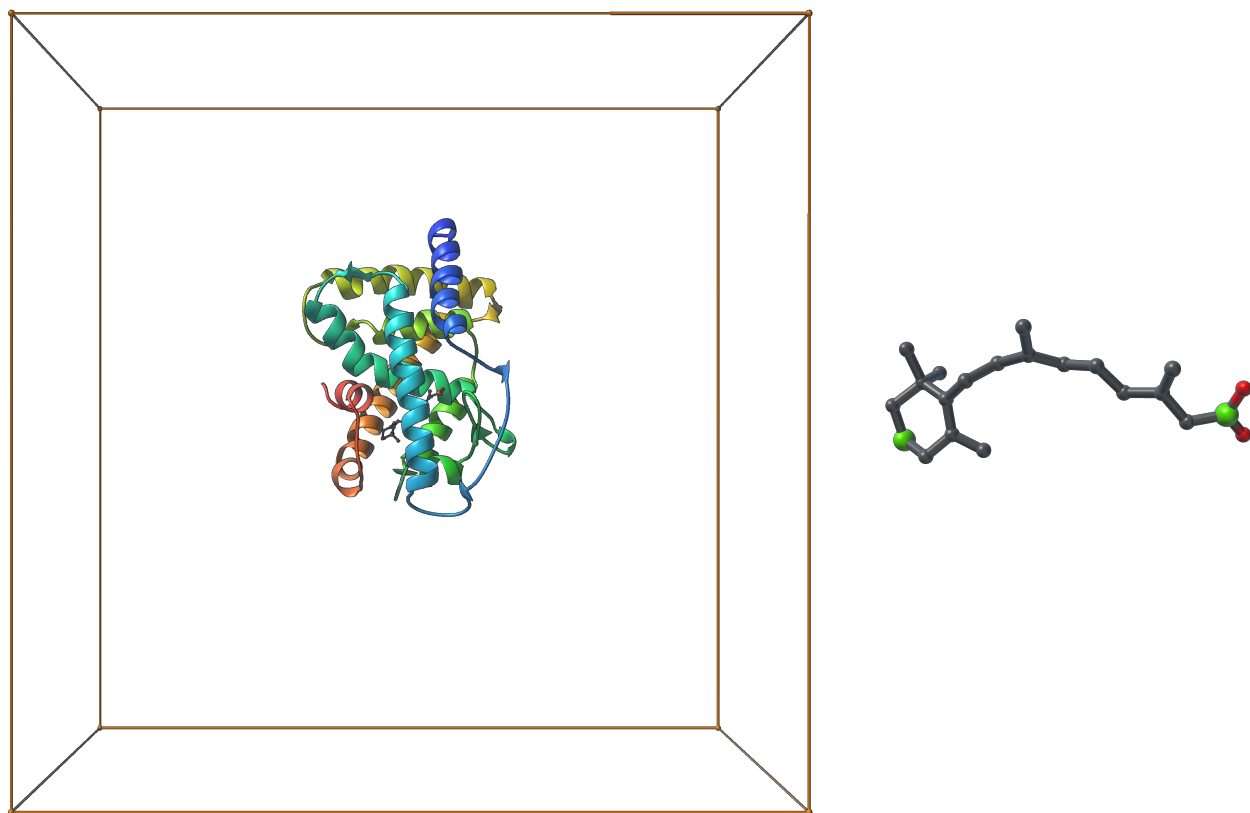


Figure 17: Left: the system (in ribbons) inside a cubic sampling region for active atoms in the ligand of Benchmark III. Right: a closer view on the ligand with active atoms in green.

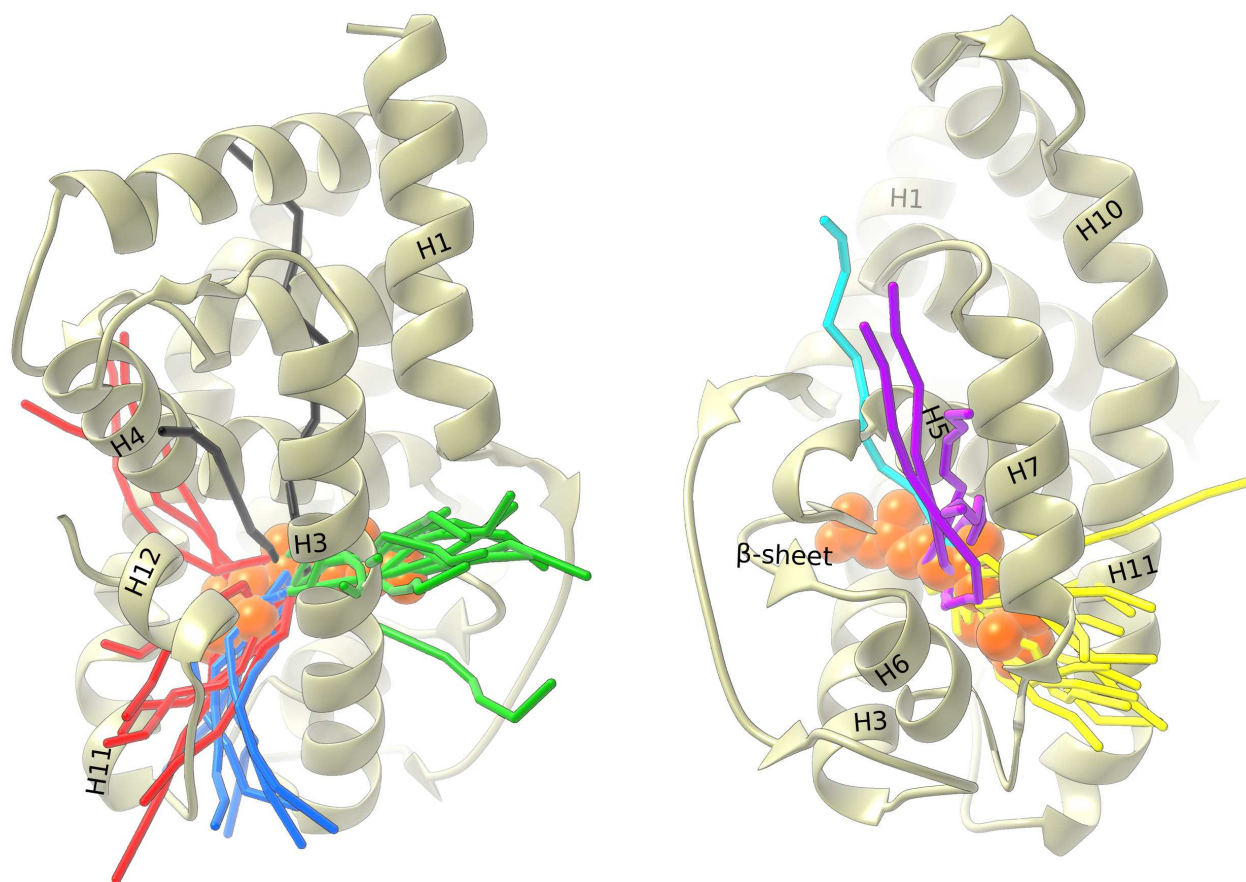


Figure 18: Paths (in colored sticks) obtained by ART-RRT for Benchmark III. The protein is represented by ribbons and the ligand by orange balls. Two different views are shown for clarity. The left picture shows pathways I in red, II in blue, III in green and Other in black. The right picture shows pathways IV in yellow, V in purple and VI in cyan.



## References

1. H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, et al., *Acta Crystallographica Section D* **58**, 899 (2002).
2. E. Yuriev and P. A. Ramsland, *Journal of Molecular Recognition* **26**, 215 (2013).
3. D. T.-H. Chang, C.-H. Ke, J.-H. Lin, and J.-H. Chiang, *Bioinformatics* **28**, 2162 (2012).
4. A. M. Gallina, P. Bisignano, M. Bergamino, and D. Bordo, *Bioinformatics* **29**, 395 (2012).
5. J. C. Fuller, M. Martinez, S. Henrich, A. Stank, S. Richter, and R. C. Wade, *Bioinformatics* **31**, 1147 (2014).
6. S. Aci-Sèche, M. Genest, and N. Garnier, *FEBS letters* **585**, 2599 (2011).
7. V. Limongelli, M. Bonomi, and M. Parrinello, *Proceedings of the National Academy of Sciences* **110**, 6358 (2013).
8. L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, et al., *Journal of the American Chemical Society* **137**, 2695 (2015).
9. P. Tiwary, V. Limongelli, M. Salvalaglio, and M. Parrinello, *Proceedings of the National Academy of Sciences* **112**, E386 (2015).
10. Y. Wang, J. M. Martins, and K. Lindorff-Larsen, *Chemical Science* (2017).
11. M. Ø. Jensen, Y. Yin, E. Tajkhorshid, and K. Schulten, *Biophysical journal* **93**, 92 (2007).
12. N. Stanley, L. Pardo, and G. De Fabritiis, *Scientific reports* **6** (2016).
13. L. Martínez, M. T. Sonoda, P. Webb, J. D. Baxter, M. S. Skaf, and I. Polikarpov, *Biophysical journal* **89**, 2011 (2005).

14. J. Shen, W. Li, G. Liu, Y. Tang, and H. Jiang, The Journal of Physical Chemistry B **113**, 10436 (2009).
15. L.-J. Yang, J. Zou, H.-Z. Xie, L.-L. Li, Y.-Q. Wei, and S.-Y. Yang, PLoS One **4**, e8470 (2009).
16. M. Peräkylä, European Biophysics Journal **38**, 185 (2009).
17. G. S. Chen and J.-W. Chern, *Computer-aided drug design* (Wiley, NY, USA, 2007).
18. C. Dellago and P. G. Bolhuis, *Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009), pp. 167–233, ISBN 978-3-540-87706-6.
19. B. J. Alder and T. E. Wainwright, The Journal of Chemical Physics **31**, 459 (1959).
20. N. Metropolis and S. Ulam, Journal of the American statistical association **44**, 335 (1949).
21. M. R. Sorensen and A. F. Voter, The Journal of Chemical Physics **112**, 9599 (2000).
22. Y. Sugita and Y. Okamoto, Chemical physics letters **314**, 141 (1999).
23. S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, et al., science **220**, 671 (1983).
24. A. Barducci, M. Bonomi, and M. Parrinello, Wiley Interdisciplinary Reviews: Computational Molecular Science **1**, 826 (2011).
25. G. M. Torrie and J. P. Valleau, Journal of Computational Physics **23**, 187 (1977).
26. A. F. Voter, Physical Review Letters **78**, 3908 (1997).
27. T. Huber, A. E. Torda, and W. F. van Gunsteren, Journal of computer-aided molecular design **8**, 695 (1994).
28. H. Grubmüller, Physical Review E **52**, 2893 (1995).
29. J. Schlitter, M. Engels, P. Krüger, E. Jacoby, and A. Wollmer, Molecular Simulation **10**, 291 (1993).

30. S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten, in *Computational molecular dynamics: challenges, methods, ideas* (Springer, 1999), pp. 39–65.
31. S. K. Lüdemann, V. Lounnas, and R. C. Wade, *Journal of molecular biology* **303**, 797 (2000).
32. P. Carlsson, S. Burendahl, and L. Nilsson, *Biophysical journal* **91**, 3151 (2006).
33. D. Kosztin, S. Izrailev, and K. Schulten, *Biophysical journal* **76**, 188 (1999).
34. A. M. Capelli and G. Costantino, *Journal of chemical information and modeling* **54**, 3124 (2014).
35. H. Jin, J. Zhu, Y. Dong, and W. Han, *RSC Advances* **6**, 10987 (2016).
36. A. Cuzzolin, M. Sturlese, G. Deganutti, V. Salmaso, D. Sabbadin, A. Ciancetta, and S. Moro, *Journal of chemical information and modeling* **56**, 687 (2016).
37. A. Dickson and S. D. Lotz, *Biophysical Journal* **112**, 620 (2017).
38. N. M. Amato and G. Song, *Journal of Computational Biology* **9**, 149 (2002).
39. L. Jaillet, F. J. Corcho, J.-J. Pérez, and J. Cortés, *Journal of computational chemistry* **32**, 3464 (2011).
40. D. Devaurs, L. Bouard, M. Vaisset, C. Zanon, I. Al-Blawi, R. Iehl, T. Siméon, and J. Cortés, *Nucleic acids research* **41**, W297 (2013).
41. J. Cortés, D. T. Le, R. Iehl, and T. Siméon, *Physical Chemistry Chemical Physics* **12**, 8268 (2010).
42. S. M. LaValle and J. J. Kuffner Jr (2000).
43. M. Rehan, M. A. Beg, S. Parveen, G. A. Damanhour, and G. F. Zaher, *PloS one* **9**, e109705 (2014).

44. M. S. Jamal, S. Parveen, M. A. Beg, M. Suhail, A. G. Chaudhary, G. A. Damanhour, A. M. Abuzenadah, and M. Rehan, PloS one **9**, e87309 (2014).
45. M. K. Nguyen, L. Jaillet, and S. Redon, Journal of Computer-Aided Molecular Design pp. 1–15 (2017).
46. H. Jónsson, G. Mills, and K. W. Jacobsen (1998).
47. E. Weinan, W. Ren, and E. Vanden-Eijnden, Physical Review B **66**, 052301 (2002).
48. C. Dellago, Free Energy Calculations pp. 249–276 (2007).
49. C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, The Journal of Chemical Physics **108**, 1964 (1998).
50. I. Al-Blawi, M. Vaisset, T. Siméon, and J. Cortés, BMC structural biology **13**, S2 (2013).
51. M. Alexa, D. Cohen-Or, and D. Levin, in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (ACM Press/Addison-Wesley Publishing Co., 2000), pp. 157–164.
52. O. Sorkine and M. Alexa, in *Symposium on Geometry processing* (2007), vol. 4.
53. A. Cuno, C. Esperança, A. Oliveira, and P. R. Cavalcanti, in *Proceedings of the 27th computer graphics international conference* (2007), pp. 115–122.
54. P. Borosán, R. Howard, S. Zhang, and A. Nealen, in *Eurographics (short papers)* (2010), pp. 41–44.
55. M. Zollhöfer, E. Sert, G. Greiner, and J. Süßmuth, in *Eurographics (Short Papers)* (2012), pp. 85–88.
56. I. Chao, U. Pinkall, P. Sanan, and P. Schröder, in *ACM Transactions on Graphics (TOG)* (ACM, 2010), vol. 29, p. 38.
57. Y.-S. Liu, H.-B. Yan, and R. R. Martin, Journal of Computer Science and Technology **26**, 548 (2011).

58. Z. Levi and C. Gotsman, Visualization and Computer Graphics, IEEE Transactions on **21**, 264 (2015).
59. INRIA., *SAMSON: Software for Adaptive Modeling and Simulation Of Nanosystems. Version 0.6.0.* (2017), URL <https://www.samson-connect.net>.
60. E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, Physical review letters **97**, 170201 (2006).
61. D. Devaurs, T. Siméon, and J. Cortés, in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on* (IEEE, 2014), pp. 2991–2996.
62. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, The journal of chemical physics **21**, 1087 (1953).
63. K. Binder, Quantum Monte Carlo Methods p. 241 (1987).
64. W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. Scott, and I. G. Tironi (1996).
65. X. Daura, A. E. Mark, and W. F. Van Gunsteren, Journal of computational chemistry **19**, 535 (1998).
66. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1**, 19 (2015).
67. A. Šali and T. L. Blundell, Journal of molecular biology **234**, 779 (1993).
68. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, Journal of computational chemistry **25**, 1605 (2004).
69. N. Guex and M. C. Peitsch, electrophoresis **18**, 2714 (1997).
70. A. W. Schttelkopf and D. M. F. van Aalten, Acta Crystallographica Section D Biological Crystallography **60**, 1355 (2004).
71. P. Cohen, Nature reviews Drug discovery **1**, 309 (2002).

- 72. S. R. Hubbard, *Nature Reviews Molecular Cell Biology* **5**, 464 (2004).
- 73. J. Zou, Y.-D. Wang, F.-X. Ma, M.-L. Xiang, B. Shi, Y.-Q. Wei, and S.-Y. Yang, *Proteins: Structure, Function, and Bioinformatics* **72**, 323 (2008).
- 74. H. Sun, S. Tian, S. Zhou, Y. Li, D. Li, L. Xu, M. Shen, P. Pan, and T. Hou, *Scientific reports* **5** (2015).
- 75. J. Abramson, I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, and S. Iwata, *Science* **301**, 610 (2003).
- 76. G. Henkelman, B. P. Uberuaga, and H. Jónsson, *The Journal of chemical physics* **113**, 9901 (2000).
- 77. G. Henkelman and H. Jónsson, *The Journal of chemical physics* **113**, 9978 (2000).
- 78. Y. Zhou, L. Guan, J. A. Freites, and H. R. Kaback, *Proceedings of the National Academy of Sciences* **105**, 3774 (2008).